

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 13, 2004, 14:35:59 ; Search time 43 Seconds

(without alignments)
803.087 Million cell updates/sec

Title: US-10-759-277-4

Perfect score: 1887

Sequence: 1 MPLELELCGRWVGQHPCHF.....EEDDTIMEELVDNHGKKIKS 359

Scoring table: BLOSUM62

Gapop 10.0 ; Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR 78:*

1: PIR1:*

2: PIR2:*

3: PIR3:*

4: PIR4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1804	95.6	359	2 JC7321	N-acetylneuraminic
2	619	32.8	337	2 H6432	Spore coat polysac
3	532.5	28.2	341	2 T44651	capsular polysacch
4	510.5	27.1	346	2 T69836	neuB protein - Esc
5	487	25.8	334	2 D81276	N-acetylneuraminic
6	419	22.2	350	2 F97169	sialic acid syntha
7	382.5	20.3	349	2 S60760	polysialic acid ca
8	370	19.6	373	2 S39722	spore coat polysac
9	358	19.0	356	2 D87604	neuB protein, prob
10	356.5	18.9	343	2 B81275	N-acetylneuraminic
11	354	18.8	343	2 H81318	N-acetylneuraminic
12	303	16.1	340	2 B64542	spore coat polysac
13	287	15.2	340	2 H71965	sialic acid syntha
14	226	12.0	378	2 H71307	spore coat polysac
15	122.5	6.5	331	2 B90173	carboxysome format
16	119	6.3	352	2 A11899	hypothetical prote
17	110	5.8	134	2 S53514	antifreeze protein
18	110	5.8	265	2 B75161	phospho-2-dehydro-
19	108.5	5.7	351	2 S74729	carboxysome format
20	108.5	5.7	633	2 D64222	DNA topoisomerase
21	108	5.7	63	2 S04974	antifreeze protein
22	108	5.7	338	2 E72388	hypothetical prote
23	107.5	5.7	337	2 A97010	daHP synthase rela
24	107	5.7	276	2 E72643	probable phospho-2
25	106.5	5.6	478	2 B70342	glutaryl-trna (Gln
26	102.5	5.4	1140	2 S73786	hypothetical prote
27	102	5.4	63	2 S04973	antifreeze protein
28	101	5.4	1107	1 S52517	myosin I heavy cha
29	101	5.4	1175	2 T20346	pyruvate carboxyla

ALIGNMENTS

RESULT 1

JC7321

N-acetylneuraminic acid 9-phosphate synthase (EC 4.1.3.-) - mouse

N:Alternate names: N-acetylneuraminic acid synthase homolog

C:Species: Mus musculus (house mouse)

C:Date: 08-Sep-2000 #sequence_revision 08-Sep-2000 #text_change 06-Oct-2000

C:Accession: JC7321

R:Nakata, D.; Close, B.E.; Colley, K.J.; Matsuda, T.; Kitajima, K.

Biochem. Biophys. Res. Commun. 273, 642-648, 2000

A:Title: Molecular cloning and expression of the mouse N-acetylneuraminic acid 9-phosphat

A:Reference number: JC7321

A:Accession: JC7321

A:Molecule type: mRNA

A:Residues: 1-359 <NAK>

A:Cross-references: DDBJ:AB041263

C:Comment: This enzyme is a cytosolic enzyme and as a housekeeping enzyme as well, which cetyl-D-mannosamine 6-phosphate to synthesize N-acetylneuraminic acid 9-phosphate.

C:Genetics:

A:Gene: sialic acid

C:Superfamily: Streptococcus agalactiae capsular polysaccharide biosynthesis protein cps

C:Keywords: carbon-carbon lyase; oxo-acid-lyase

Query Match 95.6%; Score 1804; DB 2; Length 359;
Best Local Similarity 94.2%; Pred. No. 1e-133;
Matches 338; Conservative 11; Mismatches 10; Indels 0; Gaps 0;

QY	1	MPLELELCGRWVGQHPCHIIAIGQHQGDLDAKRMIRMAKECGADCAKFKQSELEF	60
Db	1	MPLELELCGRWVGQHPCHIIAIGQHQGDLDAKRMIRMAKECGADCAKFKQSELEF	60
QY	61	KFNKALERTYTSKHSWGKTYGHEKHLEFSDQYRELQRYAEVGIFFFTASGDEWAVE	120
Db	61	KFNKALERTYTSKHSWGKTYGHEKHLEFSDQYRELQRYAEVGIFFFTASGDEWAVE	120
QY	121	FLHELVNVPFKVSGDGTNNPYLEKTAAGRPVMISSGMSMDTMKQYQIVKPLNPNFC	180
Db	121	FLHELVNVPFKVSGDGTNNPYLEKTAAGRPVMISSGMSMDTMKQYQIVKPLNPNFC	180
QY	181	FLOCTSNAYPLQPDVNLRVISEYQKLPDIPIGYSGHETGIAISVAALGAKVLERHIT	240
Db	181	FLOCTSNAYPLQPDVNLRVISEYQKLPDIPIGYSGHETGIAISVAALGAKVLERHIT	240
QY	241	LDTKWGSDHSASLEPELAEVRSVRLVERALGSPTKQLLPCEMACNEKLGSVAVKVK	300
Db	241	LDTKWGSDHSASLEPELAEVRSVRLVERALGSPTKQLLPCEMACNEKLGSVAVKVK	300
QY	301	IPEGTILTMDLTVKVGEPKAYPPEDIFNLVGGKVLVTVEEDDTIMEELVDNHGKKIKS	359
Db	301	IPAGTTTLTDLTVKVGEPKAYPPEDIFNLVGGKVLVTVEEDDTIMEELVDNHGKKIKS	359

RESULT 2

H64432
 spore coat polysaccharide biosynthesis protein E homolog - Methanococcus jannaschii
 C:Species: Methanococcus jannaschii
 C>Date: 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change 21-Jul-2000
 C:Accession: H64432
 R:Bult, C.J.; White, O.; Olsen, G.J.; Zhou, L.; Fleischmann, R.D.; Sutton, G.G.; Blake, R.; Reich, C.I.; Overbeek, R.; Kirkness, E.F.; Weinstock, K.G.; Merrick, J.M.; Glodek, A.; Aron, J.D.; Sadow, P.W.; Hanna, M.C.; Cotton, M.D.; Roberts, K.M.; Hurst, M.A.
 Science 273, 1058-1073, 1996
 A:Authors: Kaine, B.P.; Borodovsky, M.; Klenk, H.P.; Fraser, C.M.; Smith, H.O.; Woese, C.
 A:Title: Complete genome sequence of the methanogenic archaeon, Methanococcus jannaschii
 A:Reference number: A64300; MUID:96337999; PMID:8688087
 A:Accession: H64432
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-337 <BUL>
 A:Cross-references: GB:U67549; GB:L77117; NID:G2826363; PIDN:AAB99068.1; PID:gi591717; I
 C:Genetics:
 A:Map position: REV1006919-1005906
 C:Superfamily: Streptococcus agalactiae capsular polysaccharide biosynthesis protein cpe

Query Match 32.8%; Score 619; DB 2; Length 337;
 Best Local Similarity 40.0%; Pred. No. 7.3e-41;
 Matches 136; Conservative 65; Mismatches 115; Indels 24; Gaps 7;

QY 11 RWVGQHPCTIIAIGQNHQGLDVAKRMIRMAKCGADCAKFOKSELEFKNRKALEP 70
 DB 9 RYVGKEPTIIAEGNLNHDIDIGKELVKEAKCGADAIKPSVHT-----D 58
 QY 71 YTSKSWGTYGHEKHLRFSDQYRELQRYAEVGIFFTAGMDMAVEFLHNLNVPFP 130
 DB 59 FISKKS---EYELFKSLSEEEFYELKEAIGIMISPLDLKYVDILNKNVPAP 115
 QY 131 KVGSGDTNFPYLEKTAKGRPMVSSGMQMDTKQVYQIVKPLNPF-----FLOCTS 186
 DB 116 KIASGLTYPLEKVAKTGKPVILSTGMSDIG---EWEAVKVLNNGCRDILLHCIS 172
 QY 187 AYPQLQEDNLRVISEYQKLPDPIDIGYSGHETGIAISVAALGAKVLERHTLDKTKW 246
 DB 173 SYPTPYEDVNLNAIKTISIF-NIPVGYSDHTLILAPVVSVALGADVIEKHFTLDKNME 231
 QY 247 GSDHSLSEGLAEILVRSVRLVERALGSPTKQLLPCMACNEKLGKSVVAKVPIEGII 306
 DB 232 GPDHLSADPEEFKEWNNIRLVKMLGSGEKTIPMSERDVIVEARSIKVENIKKGEY 291
 QY 307 LTMDMLTVKVGEP-KAYPPEDIFNLVGKKVLTVEEDDTI 345
 DB 292 LSVDNISFK--RPGRIETKYLIIILNRKINKDKEEDII 329

RESULT 3
 T44651
 capsular polysaccharide biosynthesis protein cpsN [imported] - Streptococcus agalactiae
 C:Species: Streptococcus agalactiae
 C>Date: 21-Jan-2000 #sequence_revision 21-Jan-2000 #text_change 09-Jun-2000
 R:Chaffin, D.O.; Yim, H.H.; Beres, S.B.; Sweet, E.S.; Nittayajarn, A.; Rubens, C.E.
 submitted to the EMBL Data Library, June 1999
 A:Reference number: Z22821
 A:Accession: T44651
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-341 <CHA>
 A:Cross-references: EMBL:AF163833; PIDN:AAD53074.1
 A:Experimental source: strain COH1; serotype III
 C:Genetics:
 A:Gene: cpsN
 C:Superfamily: Streptococcus agalactiae capsular polysaccharide biosynthesis protein cpe

Query Match 20.2%; Score 532.5; DB 2; Length 341;
 Best Local Similarity 36.8%; Pred. No. 4.5e-34;
 Matches 125; Conservative 66; Mismatches 136; Indels 13; Gaps 6;

QY 20 FTIAEIQNHQGLDVAKRMIRMAKCGADCAKFOKSELEFKNRKALEPPTSKHSWK 79
 DB 3 YIIAEIGCNHNGDINLAKWVDVAVSCGVDAVQFOTFAEKALISLKFAPKAEY-OKATTGT 61
 QY 80 TYG--EHRKHLEFSDQYRELQRYAEVGIFFTAGMDMAVEFLHNLNVPFPKVGSDT 137
 DB 62 ADSQLEMTKRLSESPYELMRDYAISKGVETFTFPDEESLEFLISTDMPIYKIPSGBI 121
 QY 138 NNFPYLEKTAKGRPMVSSGMQMDTKQVYQIVKPLNPN---FCFLQCTSAYPELOPE 193
 DB 122 TNLPLEKIGKQKKVILSTGMA---VMEELHQAQVNLRLQNGTTDISILHCTEYPTFP 178
 QY 194 DVNLRVISEYQKLPDPIDIGYSGHETGIAISVAALGAKVLERHTLDKTKWGS DHSAS 253
 DB 179 SLNLNVIHTLKDEFKDLTIGYSDHSIGSEVPITAAAMGAETIEKHFTLDTNMEGPDHKA 238
 QY 254 LEPGELAEILVRSVRLVERALGSPTKQLLPCMACNEKLGKSVVAKVPIEGTILTM DMLT 313
 DB 239 ATPDILAAALVKGVRIVEQALGRFEKIPDFVEEKNKIVARKSVVALPKPKKGDYIS ENIT 298
 QY 314 KVGGEPP-KAYPPEDIFNLVGKKVLTVEEDDTIMEELVDN 352
 DB 299 VK--RPGNGISPMWYDILGQEAQDDFEDEVIRDSRFPEN 336

RESULT 4
 I69836
 neuB protein - Escherichia coli
 C:Species: Escherichia coli
 C>Date: 07-Jun-1996 #sequence_revision 07-Jun-1996 #text_change 20-Jun-2000
 C:Accession: I69836
 R:Annunziato, P.W.; Wright, L.F.; Vann, W.F.; Silver, R.P.
 J. Bacteriol. 177, 312-319, 1995
 A:Title: Nucleotide sequence and genetic analysis of the neuD and neuB genes in region 2
 A:Reference number: I55145; MUID:95113767; PMID:7814319
 A:Accession: I69836
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-346 <RES>
 A:Cross-references: EMBL:U05248; NID:G454079; PIDN:AAC43302.1; PID:G454081
 C:Superfamily: Streptococcus agalactiae capsular polysaccharide biosynthesis protein cpe

Query Match 27.1%; Score 510.5; DB 2; Length 346;
 Best Local Similarity 36.3%; Pred. No. 2.4e-32;
 Matches 123; Conservative 61; Mismatches 140; Indels 15; Gaps 6;

QY 20 FTIAEIQNHQGLDVAKRMIRMAKCGADCAKFOKSELEFKNRKALEPPTSKHSWG- 78
 DB 5 YIVAEIGCNHNGSDVIAREMILKAKBAGVNAVQFTFKADKLISAIAPKAEYQIKNGEL 64
 QY 79 KTYGHEKHLRFSDQYRELQRYAEVGIFFTAGMDMAVEFLHNLNVPFPKVGSDTN 138
 DB 65 ESQLEMTKLEKDYDLHLMYAVSLNDLVFTFPDEDSIDFLASLKQIKWIPSGELL 124
 QY 139 NNFPYLEKTAKGRPMVSSGMQMDTKQVYQIV---KPLNPFCLQCTSAYPELOP 192
 DB 125 NLPYLEKIAKLPIDPKKIISTGMATIDEIKQSVFIPNNKVPVGNITILHCTEYPTFP 184
 QY 193 EDNLRVISEYQKLPDPIDIGYSGHETGIAISVAALGAKVLERHTLDKTKWGS DHS 252
 DB 185 EDVNLAINDLKKHFFPKNNIGFSDHSGGFYAAIAAPVPGITFTIEKHFTLDKSGPDHLA 244
 QY 253 SLEPGELAEILVRSVRLVERALGSPTKQLLPCMACNEKLGKSVVAKVPIEGTILTM DML 312
 DB 245 SIPEDELKHLICIGRCVKEKSLGSKSVVTASERKNKIVARKSIITAKTEIKKGEVSEKNI 304
 QY 313 TVKVGEP-KAYPPEDIFNLVGKKVLTVEEDDTIMEELV 350
 DB 305 TTK--RPGNGISPMWYDILGK-----IABQDFIPDELI 336

RESULT 5
 D81276

N-acetylneuraminic acid synthetase (EC 4.1.1.3.-) Cj1327 [imported] - Campylobacter jejuni

C:Species: Campylobacter jejuni

C>Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 03-Jun-2002

C:Accession: D81276

R:Parkhill, J.; Wren, B.W.; Mungall, K.; Ketley, J.M.; Churcher, C.; Basham, D.; Chillingworth, C.W.; Quail, M.; Rajandream, M.A.; Rutherford, K.M.; VanVleet, A.; Whitehead, S.; Barrall, N.

A:Title: The genome sequence of the food-borne pathogen Campylobacter jejuni reveals hyp Nature 403, 665-668, 2000

A:Reference number: A81250; MUID:20150912; PMID:10688204

A:Accession: D81276

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-334 <PAR>

A:Cross-references: GB:AL119078; GB:AL111168; NID:G6968723; PIDN:CAB73754.1; PID:G6968723

A:Experimental source: serotype O2, strain NCTC 11168

C:Genetics:

A:Gene: neuB2; Cj1327

C:Superfamily: Streptococcus agalactiae capsular polysaccharide biosynthesis protein cps

C:Keywords: carbon-carbon lyase; oxo-acid-lyase

Query Match 25.8%; Score 487; DB 2; Length 334;
Best Local Similarity 35.1%; Pred. No. 1.6e-30;
Matches 118; Conservative 64; Mismatches 138; Indels 16; Gaps 5;

Qy 21 IIAIGQNHOGDLDVAKMTMAKECGADCAKQKSELEFKNRKALERYPTSKHSG-K 79
Db 6 IIAEAGVNHNGDLDLAKLIEIAADSGDFVKQSFKAKNKICSTKAKAPYQLTKTANDE 65
Qy 80 TYGEHKRHLFEFSDQVRELQRYAEVGIFFTASGMDMAVEFLHNLVPPFKVSGDNTN 139
Db 66 SLOQMVQKLELDLKAHKLILHAKKNIAFLSTPFDLESVDLLNELGLTKIFKIPSGEITN 125
Qy 140 FPLYEKTAKGRPMWISSGQMSMDTKQVQIV---KPLNPNFCFLQCTSAAYPLQPEDVN 196
Db 126 LPYLKTIKLNKKIILSTGMANLGEIEEALNVLCKKCAKQNTLLHCTEYAPFNEVN 185
Qy 197 LRVISEYQKLPDPIQYSGHETGIAISVAALGAKVLERHTLDTKWGSDHSLAP 256
Db 186 LKAMQSLKDAF-KLDVGSYDHTRGIIHSLAAVALGACVIEKHFTLDKNWSGPDHKASLEP 244
Qy 257 GEIAELVRSVRLVERALGSPTKQLLPEMACNEKLGKSVVAKVPIEGTILTMDMLTVKV 316
Db 245 QELKMLCTQIRQKAWGDGIGKASKSEQKNINVRKSLVAKKDIKKEIFSEGNLTTK- 303
Qy 317 GEPRKAYPPEDIF-----NLVGKKVLVTVEDDTIME 347
Db 304 -----RPANGISAMRYEEFLGKATKNYKEDELIRE 334

RESULT 6

F97169

sialic acid synthase [imported] - Clostridium acetobutylicum

C:Species: Clostridium acetobutylicum

C>Date: 14-Sep-2001 #sequence_revision 14-Sep-2001 #text_change 30-Sep-2001

C:Accession: F97169

R:Molling, J.; Breton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; Lee, J.; Daly, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.

J. Bacteriol. 183, 4823-4838, 2001

A:Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium Clostridium acetobutylicum

A:Reference number: A96900; MUID:21359325; PMID:21359325

A:Accession: F97169

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-350 <KUR>

A:Cross-references: GB:AE001437; PIDN:AAK80145.1; PID:G15025183; GSPDB:GN00168

A:Experimental source: Clostridium acetobutylicum ATCC824

C:Genetics:

A:Gene: CAC2187

C:Superfamily: Streptococcus agalactiae capsular polysaccharide biosynthesis protein cps

Query Match 22.2%; Score 419; DB 2; Length 350;
Best Local Similarity 32.8%; Pred. No. 3.6e-25;
Matches 112; Conservative 73; Mismatches 132; Indels 30; Gaps 10;

A;Title: The genome sequence of the food-borne pathogen Campylobacter jejuni reveals hyp
A;Reference number: A81250; MUID:20150912; PMID:10688204

A;Accession: B81275
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-343 <PAR>
A;Cross-references: GB:AL139078; GB:AL111168; NID:G6968723; PIDN:CAB73744.1; PID:G696875
A;Experimental source: serotype O2, strain NCTC 11168
C;Genetics:
A;Gene: neuB3; Cj1317
C;Superfamily: Streptococcus agalactiae capsular polysaccharide biosynthesis protein cps
C;Keywords: carbon-carbon lyase; oxo-acid-lyase

Query Match 18.9%; Score 356.5; DB 2; Length 343;
Best Local Similarity 29.9%; Pred. No. 2.8e-20;
Matches 104; Conservative 65; Mismatches 138; Indels 41; Gaps 10;

Qy 20 FIIAIEIGNHOGDLDVAKRMIRMAKCEGADCAKFO-----KSELEFK---PNR 64

Db 13 FIIAELSANHAGSLEMAKLSIAKAKKAGADAIAIKTYTPDSLTLSNKKEDFIIGGLWDK 72

Qy 65 KALERPPTS---KHSWGKTYGEHKKHLEFSDHQVRELQRYAEVGIFFTAGMDMAVEF 121

Db 73 RKLYELYESAKTPYEW-----HSQIFET---AQNEGILCFSSPPFAKEDVEF 115

Qy 122 LHELNVPPFKVSGDTNFPYLEKTAKGRPMVSSGQSDMTWKQVQIVK-PLNPNFC 180

Db 116 LKRPDPTAYKIASPEANDENFVRIAEKKFTVSTGIATEEELFKICEIFKEKNPDV 175

Qy 181 FLOCTSAYPELQPEVDNLRVISEYOKLPDIPIGYSGHETGIAISVAALCAKVLHERHIT 240

Db 176 FLKCTSTPYTAIEDMNLKGVLSLKEF-NVEVGLSDHSFGFLAPMAVALGARVIEKHF 234

Qy 241 LDKTWGSDSHASLEPEGLAELVRSVLVERALGSPTKQLLPCMACNEKLGKSVARVK 300

Db 235 LDKSIESEDSKFSDFDEFKAMVDVAVQAESALGD-GKLDLDEKVLKRVFARSLYASKD 293

Qy 301 IPEGTITLMDLMTYKVGEPK-AYPEPDEFNLVGVKLVTVVEEDTIME 347

Db 294 IKKGEMFSEE--NVKSVRPSFGLHPKPYQELLGKKASKDKIKFGDALIKQ 339

Query Match 16.1%; Score 303; DB 2; Length 340;
Best Local Similarity 27.7%; Pred. No. 4.3e-16;
Matches 91; Conservative 66; Mismatches 149; Indels 22; Gaps 10;

Qy 18 PCFIIAIEIGNHOGDLDVAKRMIRMAKCEGADCAKFOK-----SELBEKFNKALERPPT 72

Db 4 PPKIVAEISANHQDLNLAKESLHAIKESGADFVKLQTYTPSCWTLNSKEDPFIQGTW 63

Qy 73 SKHSWGKTYGEHKKHLEFSDHQVRELQRYAEVGIFFTAGMDMAVEFLHNLVPPFKV 132

Db 64 DKNLYELYOKASTPLEW-HAELFELARKL-DLGIF--SSPFSQAELLESLNCPMYKI 119

Qy 133 GSGDTNFPYLEKTAKGRPMVSSGQSDMTWKQVQIVKPLNPNF--CFLOCTSAYPEL 190

Db 120 ASPEIVDLDLIEKAARTQKPIILSSGIATGATEQDAISLCRRVN-NFDITLLKCVSAYS 178

Qy 191 QPEDVNLRVISEYOKLPDIPIGYSGHETGIAISVAALCAKVLHERHITDKTWGSDH 250

Db 179 KIEDANLLSMVGLGEIF-GVKFGLSDHTIGSLCPILATTIGASWIEKHFLINKSLQTPDS 237

Qy 251 SASLEPEGLAELVRSVLVERALGSPTKQLLPCMACNEKLGKSVAKVKEPIGTILT-- 308

Db 238 AFSDMFNGFKSMVEAIKQSVLALGEEPRINPKLEKRRFFARSLFVKIDIQKGEALTEN 297

Qy 309 -MDMLTVKVG-EPKAYPPEDIFNLVGVK 334

Db 298 NIKALRPNLGLHPEFKYK-----EILGQK 320

Query Match 18.8%; Score 354; DB 2; Length 343;
Best Local Similarity 30.5%; Pred. No. 4.4e-20;
Matches 96; Conservative 57; Mismatches 156; Indels 6; Gaps 3;

Qy 21 IIAIEIGNHOGDLDVAKRMIRMAKCEGADCAKFOKSELEFKFNKALE-RPYTSKHSWGK 79

Db 17 IIEIIGINHNGSLIEIAKMAKRAKAKIHKQTHIVEDEMSQEAKNVPGNANIS--- 73

Qy 80 TYGEHKKHLEFSDHQVRELQRYAEVGIFFTAGMDMAVEFLHNLVPPFKVSGDTN 139

Db 74 -IYEIMEQCALNYKDELALKYVEKQGLVYLSTPESRAAANRLEDMGVSAVYIGSGECNN 132

Qy 140 FPYLEKTAKGRPMVSSGQSDMTWKQVQIVKPLNPNFCFLOCTSAYPELQPEVDNLRV 199

Db 133 YPLIKHIAQFKPMIISIGMNSIESIKPTVKILRDYEIFPVLHTTNLYPTPSHLVRLQA 192

Qy 200 ISEYOKLPDIPIGYSGHETGIAISVAALCAKVLHERHITLDKTWGSDSHASLEPEGL 259

Db 193 MLELYKEP-NCLYGLSDHTTNLACIGAIALGASVLERHFTDMDRKGPDI VCSMDDESTL 251

Qy 260 AELVRSVLVERALGSPTKQLLPCMACNEKLGKSVAKVKEPIGTILTMTMLTVKVGEP 319

Db 252 KDLINQTOEVLVLLGDNKNKPLKEEQVTIDFAFASVWSIKDKIKKGEILSMNDNIWVRPSK 311

Qy 320 KAYPPEDIFNLVGVK 334

Db 312 GGISAKOFEAILGKR 326

RESULT 12

B64542

spore coat polysaccharide biosynthesis protein E - Helicobacter pylori (strain 26695)

C;Species: Helicobacter pylori

C;Date: 09-Aug-1997 #sequence_revision 09-Aug-1997 #text_change 20-Jun-2000

C;Accession: B64542

R;Tomb, J.F.; White, O.; Kerlavage, A.R.; Clayton, R.A.; Sutton, G.G.; Fleischmann, R.D.;

Peterson, S.; Loftus, B.; Richardson, D.; Dodson, R.; Khalak, H.G.; Glodek, A.; McKenney

Nature 388, 539-547, 1997

A;Authors: Wallin, E.; Hayes, W.S.; Borodovsky, M.; Karpk, P.D.; Smith, H.O.; Fraser, C.N

A;Title: The complete genome sequence of the gastric pathogen Helicobacter pylori.

A;Reference number: A64520; MUID:97394467; PMID:9252185

A;Accession: B64542

A;Status: preliminary; nucleic acid sequence not shown; translation not shown

A;Molecule type: DNA

A;Residues: 1-340 <TON>

A;Cross-references: GB:AB000538; GB:AB000511; NID:G2313263; PIDN:AAD07248.1; PID:G2313267

C;Superfamily: Streptococcus agalactiae capsular polysaccharide biosynthesis protein cps

RESULT 13

B71965

sialic acid synthase - Helicobacter pylori (strain J99)

C;Species: Helicobacter pylori

A: Variety: strain J99
C: Date: 12-Feb-1999 #sequence_revision 12-Feb-1999 #text_change 20-Jun-2000
C: Accession: B71965
R: Alm, R.A.; Ling, L.S.L.; Moir, D.T.; King, B.L.; Brown, E.D.; Doig, P.C.; Smith, D.R.;
; Ives, C.; Gibson, R.; Merberg, D.; Mills, S.D.; Jiang, Q.; Taylor, D.E.; Vovis, G.F.;
Nature 397, 176-180, 1999
A: Title: Genomic sequence comparison of two unrelated isolates of the human gastric path
A: Reference number: A71800; MUID: 99120557; PMID: 9923682
A: Accession: B71965
A: Status: preliminary
A: Molecule type: DNA
A: Residues: 1-340 <ARN>
A: Cross-references: GB:AE001455; GB:AE001439; NID: g4154678; PIDN: AAD05747.1; PID: g415468
A: Experimental source: strain J99
C: Genetics:
C: Superfamily: Streptococcus agalactiae capsular polysaccharide biosynthesis protein cps

Query Match 15.2%; Score 287; DB 2; Length 340;
Best Local Similarity 27.8%; Pred. No. 7.8e-15;
Matches 92; Conservative 68; Mismatches 137; Indels 34; Gaps 14;

QY 21 IIAEIQNHQGGDLVAKRMIRMAKCGADCAKFKQ-----SELEPKFNKALERPYSKH 75
DB 7 IVAELSANHNQDLNLAKESHAIKESGADFVKLTQVTPSCMTLDSK-----EDPFIQ 60

QY 76 S-WGK--TYGEHR---HLEFSDHQYRELQVAEVEGIFFTASGMDMAVEFLHNLVPP 129
DB 61 TLWDKENLYGLYKASTPLEW-HAELELAKKL-DLIGIF--SSPSSKALELLESLDCPM 116

QY 130 FVKGSGDTNNFPYLEKTAKKGRPMVWISSQMGMDTMKVQYQVVKPLNPNF--CFLOQTS 187
DB 117 YIAEFAEIVDLLEKAAQTQPIILSSGIATHAELODAISLCRGVN-NFDITLKCVA 175

QY 188 YLQPEDNLRVISEYQKLFDPDIPIGYSGHETGIAISVAALGAKVLERHITLQTKWG 247
DB 176 YPSKIEDAHLMSVKGTF-GVKFGLSDHTIGSLCPILATILGASMIKHFILNKSQT 234

QY 248 SHSASLEGEALVRSVRLVERALGSPTKQLLPCEMACNEKLGKSVVAKVKIPGTTIL 307
DB 235 PDSAFSMDPFGFKSMVGAIKQSVLALGEBEKPINPKTLERRFFAKSLFVFKIQKEAL 294

QY 308 TMD---MLTVKVG-EPKAYPPEDIFNLVGKK 334
DB 295 TSDNIKALPNLGLHKKFYK-----EILGQK 320

RESULT 14
H71307
probable spore coat polysaccharide biosynthesis protein (spse) - syphilis spirochete
C: Species: Treponema pallidum subsp. pallidum (syphilis spirochete)
C: Date: 24-Jul-1998 #sequence_revision 24-Jul-1998 #text_change 19-Apr-2002
C: Accession: H71307
R: Fraser, C.M.; Norris, S.J.; Weinstock, G.M.; White, O.; Sutton, G.G.; Dodson, R.; Gwin
rson, J.; Khalak, H.; Richardson, D.; Howell, J.K.; Chidambaram, M.; Utterback, T.; McDo
they, L.; Weidman, J.; Smith, H.O.; Venter, J.C.
Science 281, 375-388, 1998
A: Title: Complete genome sequence of Treponema pallidum, the syphilis spirochete.
A: Reference number: A71250; MUID: 98332770; PMID: 9665876
A: Accession: H71307
A: Status: preliminary; nucleic acid sequence not shown; translation not shown
A: Molecule type: DNA
A: Residues: 1-378 <COL>
A: Cross-references: GB:AE001232; GB:AE000520; NID: g3322856; PIDN: AAC65539.1; PID: g332285
A: Experimental source: strain Nichols
C: Genetics:
A: Gene: TP0562
C: Superfamily: Streptococcus agalactiae capsular polysaccharide biosynthesis protein cps

Query Match 12.0%; Score 226; DB 2; Length 378;
Best Local Similarity 25.1%; Pred. No. 5.4e-10;
Matches 85; Conservative 47; Mismatches 166; Indels 40; Gaps 9;

QY 8 CPGRWVGGQHPCFIITAEIQNHQGGDLVAKRMIRMAKCGADCAKFKQSELEFKFNKAL 67
DB 4 CGGRCFRDPADILITAEIGSAHAGSFRARALIDAAADAAAANKFKQ-----LIVAHEI 57

QY 68 ERPYSK---HSGWGTYYGEHRHLEFSDHQYRELQVAEVEGIFFTASGMDMAVEFLHE 124
DB 58 LHPLTGAVRLPSGAVSLYQRFEELEVPLSFYAQCFNHARSRLVGLISPPGPRSALEA 117

QY 125 LNVPEFKVSGDTNNFPYLEKT-AKKGRPMVWISSGM---QSMDTMKVQYQVVKPLNPNF 179
DB 118 LKPDFLKVASPEL-NYPTLITSLAAEELPLILSSGVCLLKEIEGALAQRQYTKQ-GSSH 175

QY 180 CFLQCTSAYPELQPEDNLRVISEYQKLFDPDIPIGYSGHETG-IAISVAALGAKVLERH 238
DB 176 ALLHCITAYPAPETEVNLAALLPALATIF-NINVGVDHSVDPLLVPLLARAHGACIVEKH 234

QY 239 ITLDTWKSGDHSASLEGEALVRSV-----RLVERALGSP 276
DB 235 ICLSRDTAGLDDSIADPADFTMTAALNSCARRSPSQIISFLHERGYAPHVVRAVIGSG 294

QY 277 TKQLLPCEMACNEKLGKSVVAKVKIPGTTILTMDMLTV 314
DB 295 EKVLPASERAHYQKNSRLHYLHAYPRGTVLQKENLVI 332

RESULT 15
B90173
hypothetical protein aroG [imported] - Sulfolobus solfataricus
C: Species: Sulfolobus solfataricus
C: Date: 24-May-2001 #sequence_revision 24-May-2001 #text_change 18-Nov-2002
C: Accession: B90173
R: She, Q.; Singh, R.K.; Confalonieri, F.; Zivanovic, Y.; Allard, G.; Awayez, M.J.; Chan-
Jong, I.; Jeffries, A.C.; Kozera, C.J.; Medina, N.; Peng, X.; Thi-Ngoc, H.P.; Redder, P.
arrett, R.A.; Ragan, M.A.; Sensen, C.W.; Van der Oost, J.
submitted to GenBank, April 2001
A: Description: Sulfolobus solfataricus complete genome.
A: Reference number: A99139
A: Accession: B90173
A: Status: preliminary
A: Molecule type: DNA
A: Residues: 1-331 <KUR>
A: Cross-references: GB:AE006641; NID: g13813447; PIDN: AAK40641.1; GSPDB: GN00155
C: Genetics:
A: Gene: aroG
C: Superfamily: phospho-2-dehydro-3-deoxyoctonate aldolase

Query Match 6.5%; Score 122.5; DB 2; Length 331;
Best Local Similarity 23.9%; Pred. No. 0.059;
Matches 66; Conservative 46; Mismatches 103; Indels 61; Gaps 15;

QY 18 PCFIITAEIQNHQGGDLVAKRMIRMAKCGADCAKFKQSELEFKFNKALERPYSKH 77
DB 95 PCAV-----ENEQVUTTKAV-----KRAGASLL-----RGKAYKRTSPYSF 133

QY 78 GKTYYGSHRHLFSDHQYRELQVAEVEGIFFTASGMDMAVEFLHNLVPPFKVSGDT 137
DB 134 -QQLGE-----EGVKILRRVGVDEVLPIVTEIMDTRDSNIFSQY-VDMIQICARNA 182

QY 138 NNFPYLEKTAKKGRPMVWISSQMGMDTMKVQYQVVKPL-----NENFC-----FLQCTS 186
DB 183 QNFSLLKEVGLGKPKVLKRGMG--NTVEEWLQAAEYILLEGNGNTVLCERGIRTFEKT 240

QY 187 AVPLQPEDNLRVISEYQKLFDPDIPY-GYSGHETG-----IAISVAALGAKVL---ERH 238
DB 241 RFTL---DIGMWAA---KLMTHLPICADPSHPAGKRELVHSLAALAAAGADMILLIEVH 294

QY 239 ITLDTWKSGDHSASLEGEALVRSVRLVERALG 274
DB 295 PHPEKAL--SDSBOQLTPESFEVLMNRIRTLARALG 328

Search completed: September 13, 2004, 14:44:25
Job time : 48 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 13, 2004, 14:24:08 ; Search time 25 Seconds
(without alignments)
747.727 Million cell updates/sec

Title: US-10-759-277-4
Perfect score: 1887
Sequence: 1 MPLELECPGRWVGQHPCF.....EEDDTIMEELVDNHGKIKS 359

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_42:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1883	99.8	359	1 SIAS_HUMAN	Qnr45 homo sapien
2	619	32.8	337	1 YA65_METJA	Q58465 methanococ
3	370	19.6	373	1 SPSE_BACSU	P39625 bacillus su
4	113.5	6.0	1248	1 SM1B_MOUSE	Q920f5 mus musculu
5	110	5.8	134	1 ANP3_RHIDE	P35753 rhigophila
6	108.5	5.7	633	1 PARE_MYCGE	P47445 mycoplasma
7	108	5.7	63	1 ANP2_AUSBR	P12101 austrolycic
8	106.5	5.6	478	1 AROF_THEMA	Q9wyh8 thermotoga
9	102.5	5.4	1140	1 GATB_AQUAE	O66766 aquifex aeo
10	102	5.4	63	1 YD76_MYCPN	P75405 mycoplasma
11	102	5.4	63	1 ANP1_AUSBR	P12100 austrolycic
12	101	5.4	1107	1 MY1E_RAT	Q63356 rattus norv
13	100.5	5.3	469	1 EGLI_BACSU	P40740 bacillus su
14	100	5.3	1109	1 MY1E_HUMAN	Q12965 homo sapien
15	98.5	5.2	371	1 GPR_BACME	P23321 bacillus me
16	97.5	5.2	642	1 DNAK_XANCP	O8pak9 xanthomonas
17	97	5.1	378	1 ARGE_VIBPA	P59601 vibrio para
18	96.5	5.1	641	1 DNAK_XANAC	Q8pm50 xanthomonas
19	96.5	5.1	1437	1 DPO3_LACPL	Q88vk2 lactobacill
20	96	5.1	880	1 FODK_RICPR	Q9zd55 rickettsia
21	95.5	5.1	513	1 CHLB_MARPO	P26238 marchantia
22	95.5	5.1	2869	1 RBPI_PLAVB	O00798 plasmodium
23	94	5.0	64	1 ANP2_RHIDE	P12102 rhigophila
24	93.5	5.0	512	1 GPMI_CLOPE	Q8xku2 clostridium
25	93	4.9	84	1 ANP1_RHIDE	P12015 rhigophila
26	93	4.9	542	1 CYMO_ACTIP	P12015 acinetobact
27	92.5	4.9	1633	1 YP74_CAEEL	O99221 caenorhabdi
28	92	4.9	346	1 BPHI_BURCE	P51015 burkholderi
29	92	4.9	638	1 DNAK_XYLF	Q9pb05 xylella fas
30	92	4.9	1656	1 ATCC_YEAST	Q12674 saccharomyc
31	91.5	4.8	458	1 EF11_RHIRA	P06805 rhizomucor
32	91.5	4.8	458	1 EF12_RHIRA	P14864 rhizomucor
33	91	4.8	91	1 ANP3_MACAM	P19606 macrozoarce

RESULT 1

SIAS_HUMAN

AC Q9NR45; Q8WUV9; Q9BWS6; Q9NVD4; STANDARD; PRT; 359 AA.
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Sialic acid synthase (N-acetylneuraminase synthase) (EC 2.5.1.56) (N-acetylneuraminic acid synthase) (N-acetylneuraminic acid phosphate synthase) (EC 2.5.1.57) (N-acetylneuraminic acid phosphate synthase).
 GN SAS.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A., AND CHARACTERIZATION.
 RC TISSUE=Liver.
 RX MEDLINE=20298869; PubMed=10749855;
 RA Lawrence S.M., Huddleston K.A., Pitts L.R., Nguyen N., Lee Y.C., Vann W.F., Coleman T.A., Beterbaugh M.J.;
 RT "Cloning and expression of the human N-acetylneuraminic acid phosphate synthase gene with 2-keto-3-deoxy-D-glycero-D-galacto-nononic acid biosynthetic ability.";
 RT J. Biol. Chem. 275:17869-17877 (2000).
 [2]
 SEQUENCE FROM N.A.
 RA Isogai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y., Nishikawa T., Nagai K., Sugano S., Aotsuka S., Yoshikawa Y., Matsunawa H., Ishii S., Kawai Y., Saito K., Yamamoto J., Wakamatsu A., Nakamura Y., Nagahari K., Masuho Y., Sasaki N.;
 RA "NEDO human cDNA sequencing project.";
 RT Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
 [3]

SEQUENCE FROM N.A., AND VARIANT ASP-68.

TISSUE=Lung, and Placenta;

MEDLINE=22388257; PubMed=12477932;

RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,

Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,

Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,

Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Haieff F.,

Diatchenko L., Marusina K., Farmer A., Rubin G.M., Hong L.,

Shapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,

Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,

Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,

Bosak S.A., McSwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,

Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,

Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,

Fahey J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,

Whiting M., Madan A.C., Young A.C., Shevchenko Y., Bouffard G.G.,

Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C.,

Rodriguez A.C., Guimond J., Schmutz J., Myers R.M.,

Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smallos D.E.,

Schmerch A., Schein J.E., Jones S.J.M., Marra M.A.;
 "Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences."

Q8k990 buchnera ap
 Q95078 drosophila
 P36298 canine ente
 O8t428 methanopyru
 P75706 escherichia
 Q87bs8 xylella fas
 P45597 x multiphos
 P40151 saccharomyc
 P24028 lycodes pol
 P13522 streptococc
 Q8gjn0 synecococc
 Q9dbs9 mus musculu

ALIGNMENTS

```

RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
CC -!- FUNCTION: Produces N-acetylneuraminic acid (Neu5Ac) and 2-keto-3-
CC deoxy-D-glycero-D-galacto-nononic acid (KDN). Can also use N-
CC acetylmannosamine 6-phosphate and mannose 6-phosphate as
CC substrates to generate phosphorylated forms of Neu5Ac and KDN,
CC respectively.
CC -!- CATALYTIC ACTIVITY: Phosphoenolpyruvate + N-acetyl-D-mannosamine +
CC H(2)O = phosphate + N-acetylneuraminic acid.
CC -!- CATALYTIC ACTIVITY: Phosphoenolpyruvate + N-acetyl-D-mannosamine 6-
CC phosphate + H(2)O = N-acetylneuraminic acid 9-phosphate + phosphate.
CC -!- TISSUE SPECIFICITY: Ubiquitous.
CC -!- SIMILARITY: Contains 1 AFP-like domain.
CC -----
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF257466; AAF75261.1; -
DR EMBL; AK001659; BAA91818.1; -
DR EMBL; BC000008; AAH00008.1; -
DR EMBL; BC019315; AAH19315.1; -
DR MIM; 605202; -
DR HSSP; F19614; 9AME.
DR GO; GO:0005737; C:cytoplasm; NAS.
DR GO; GO:0008781; F:N-acetylneuraminate cytidyltransferase acti. .; NAS.
DR GO; GO:0009103; P:lipopolysaccharide biosynthesis; NAS.
DR InterPro; IPR006014; Antifreeze dom.
DR InterPro; IPR004144; Neuf.
DR Pfam; PF01354; Antifreeze; 1.
DR ProDom; PD003258; AntifreezeIII; 1.
DR PROSITE; PS0844; AFP_LIKE; 1.
DR Transferrase; Polymorphism.
FT DOMAIN 294 353 APP-LIKE.
FT VARIANT 68 68 E -> D.
FT FTID=VAR 013308.
FT CONFLICT 232 232 A -> T (IN REF. 2).
FT CONFLICT 321 321 G -> A (IN REF. 1).
FT SEQUENCE 359 AA; 40307 MW; 2E02D47F4F98592F CRC64;

Query Match 99.8%; Score 1883; DB 1; Length 359;
Best Local Similarity 99.7%; Pred. No. 7.2e-136;
Matches 358; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MPLELELCPRWVGQHPGCFIIAIGONHOGDLDVAKRMIRMAKCEGADCAKFKQSELEF 60
DB 1 MPLELELCPRWVGQHPGCFIIAIGONHOGDLDVAKRMIRMAKCEGADCAKFKQSELEF 60

QY 61 KFNKALERYPTSKHGKTYGHEKRLHFSHPQYRELQRYAEVGIFFFTASGMDMAVE 120
DB 61 KFNKALERYPTSKHGKTYGHEKRLHFSHPQYRELQRYAEVGIFFFTASGMDMAVE 120

QY 121 FLHELNVPPFKVSGDGNFPPYLEKTAAGKGRPMVVISSGMSMDTMKVQYIVKPLNPF 180
DB 121 FLHELNVPPFKVSGDGNFPPYLEKTAAGKGRPMVVISSGMSMDTMKVQYIVKPLNPF 180

QY 181 FLOCTSAIPLOPEDVNLRVISYOKLPDPIGYSGHETGTAISVAALGAKVLERHIT 240
DB 181 FLOCTSAIPLOPEDVNLRVISYOKLPDPIGYSGHETGTAISVAALGAKVLERHIT 240

QY 241 LDKTWKSDRSASLEPELAEVLRSVRLVERALGSPTKQLLPCEMAKNEKLGKSVVAKV 300
DB 241 LDKTWKSDRSASLEPELAEVLRSVRLVERALGSPTKQLLPCEMAKNEKLGKSVVAKV 300

QY 301 IPEGTILTMDLTVKVGPKAPYPEDIPNLVKKKVLVTVEBDDTIMBELVDNHHKKIKS 359
DB 301 IPEGTILTMDLTVKVGPKAPYPEDIPNLVKKKVLVTVEBDDTIMBELVDNHHKKIKS 359

```

```

RESULT 2
YAG5_METJA STANDARD; PRT; 337 AA.
AC Q58465;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Hypothetical protein MJ1065.
GN MJ1065.
OS Methanococcus jannaschii.
OC Archaea; Euryarchaeota; Methanococci; Methanococcales;
OC Methanocaldococcales; Methanocaldococcus.
OX NCBI_TaxID=2190;
RX SEQUENCE FROM N.A.
RP STRAIN=JAL-1 / DSM 2661 / ATCC 43067;
RC MEDLINE=96337999; PubMed=8688087;
RA Sutton G.G., Blake J.A., Fitzgerald L.M., Clayton R.A., Gocayne J.D.,
RA Sutton G.G., Blake J.A., Fitzgerald L.M., Adams M.D., Reich C.I.,
RA Kerlavage A.R., Dougherty B.A., Tomb J.-F., Merrick J.M., Glodek A.,
RA Overbeek R., Kirkness E.F., Weinstock K.G., Fuhrmann J.L., Nguyen D.,
RA Scott J.L., Geoghagen N.S.M., Weidman J.F., Sadow P.W., Hanna M.C.,
RA Utterback T.R., Kelley J.M., Peterson J.D., Haurst M.A., Kaine B.P., Borodovsky M.,
RA Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,
RA Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;
RA "Complete genome sequence of the methanogenic archaeon, Methanococcus
RT jannaschii";
RL Science 273:1058-1073 (1996).
CC -!- SIMILARITY: STRONG, TO B.SUBTILIS SPSE.
CC -!- SIMILARITY: Contains 1 AFP-like domain.
CC -----
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U67549; AAB99058.1; -
DR PIR; H64432; H64432.
DR TIGR; MJ1065; -
DR InterPro; IPR006014; Antifreeze dom.
DR InterPro; IPR006190; Antifreeze_like.
DR InterPro; IPR004144; Neuf.
DR Pfam; PF01354; Antifreeze; 1.
DR Pfam; PF03102; Neuf; 1.
DR PROSITE; PS0844; AFP_LIKE; 1.
DR Hypothetical protein; Complete proteome.
FT DOMAIN 279 337 AFP-LIKE.
FT SEQUENCE 337 AA; 37976 MW; 790CAFF48C111B CRC64;

Query Match 32.8%; Score 619; DB 1; Length 337;
Best Local Similarity 40.0%; Pred. No. 8.1e-40;
Matches 136; Conservative 65; Mismatches 115; Indels 24; Gaps 7;

QY 11 RWVGQHPGCFIIAIGONHOGDLDVAKRMIRMAKCEGADCAKFKQSELEFNRKALERP 70
DB 9 RYVGKGEPTIIAEGGLNHGDDIGKELVKEAKCGADAIKFSQVHTE-----D 58

QY 71 YTSKHSWGKTYGHEKRLHFSHPQYRELQRYAEVGIFFFTASGMDMAVEFLHELNVPPF 130
DB 59 FISKKS--BYVELFKSLSESEFEVELKEYAEKIGIMFISTPLDLKYVDILNKNVPAF 115

QY 131 KVSGDGNFPPYLEKTAAGKGRPMVVISSGMSMDTMKVQYIVKPLNPF 186
DB 116 KIASGDTTFYPLEKVAKTGKPVILSTGSDIG---EIVEAVKVLNNGCDRIILHCIS 172

QY 187 AYLQPEDVNLRVISYOKLPDPIGYSGHETGTAISVAALGAKVLERHITLDTKWK 246
DB 173 SYTPYEDVNLNALKTKLSIF-NIPVGYSDHTLGLAPVVSVALGADVIEKHTLDKME 231

```

QY 247 GSDHSAIEPGLAEIVRSVRLVERALGSPTRKQLLPCEMACNEKLGKSVAKVKKIPEGTI 306
 DB 232 GPDHSAIEPGLAEIVRSVRLVERALGSPTRKQLLPCEMACNEKLGKSVAKVKKIPEGTI 291
 QY 307 LTMVLTWKVGP-KAYPEPDIENLVGKVLVTVEEDDTI 345
 DB 292 LSVNDNISFK-RPGRGIETKVLSTIILNRKIKNKEEDDII 329

RESULT 3
 ID SPSE BACSU STANDARD; PRT; 373 AA.
 AC P39625;
 DT 01-FEB-1995 (Rel. 31, Last Created)
 DT 01-FEB-1995 (Rel. 31, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Spore coat polysaccharide biosynthesis protein spsE.
 GN SPSE OR IPA-67D OR BSU37870.
 OS Bacillus subtilis.
 OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
 OX NCBI_TaxID=1423;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=168;
 RX MEDLINE=95020537; PubMed=7934828;
 RA Glaser P., Kunst F., Arnaut M., Coudart M.P., Gonzales W.,
 RA Hullo M.F., Ionescu M., Lubochinsky B., Marcelino L., Moser I.,
 RA Pressecan E., Santana M., Schneider E., Schweizer J., Vertes A.,
 RA Rapoport G., Danchin A.;
 RT "Bacillus subtilis genome project: cloning and sequencing of the 97
 RT kb region from 325 degrees to 333 degrees.";
 RL Mol. Microbiol. 10:371-384(1993).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=168;
 RX MEDLINE=96044033; PubMed=9384377;
 RA Kunst F., Ogasawara N., Moser I., Albertini A.M., Alloni G.,
 RA Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,
 RA Borrias R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,
 RA Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,
 RA Choi S.K., Codani J.J., Connerth I.F., Cummings N.J., Daniel R.A.,
 RA Denicot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emmerson P.T.,
 RA Entian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,
 RA Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,
 RA Ghim S.Y., Glaser P., Goffeau A., Golightly E.J., Grandi G.,
 RA Guisepi G., Guy B.J., Haga K., Haeht J., Harwood C.R., Henaut A.,
 RA Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,
 RA Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,
 RA Kobayashi Y., Koetter P., Koningsstein G., Krogh S., Kumano M.,
 RA Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,
 RA Kurita K., Levine A., Liu H., Masuda S., Maue C., Medigue C.,
 RA Medina N., Mellado R.P., Mizuno M., Mostl D., Nakai S., Noback M.,
 RA Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,
 RA Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,
 RA Presecan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
 RA Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadaie Y.,
 RA Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,
 RA Sekiguchi J., Sekowska A., Seror S.J., Seror P., Shin B.S., Soldo B.,
 RA Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,
 RA Takeuchi M., Tamakoshi A., Tanaka T., Terpestra P., Tognoni A.,
 RA Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,
 RA Viari A., Wambutt R., Wedler E., Wedler H., Weitzenegger T.,
 RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
 RA Yoshida K., Yoshikawa H.F., Zumbstein E., Yoshikawa H., Danchin A.;
 RT "The complete genome sequence of the Gram-positive bacterium Bacillus
 RT subtilis.";
 RL Nature 390:249-256(1997).
 CC -!- PATHWAY: Spore coat polysaccharide biosynthesis.
 CC -!- SIMILARITY: Contains 1 AFP-like domain.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its

use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).

 DR EMBL; X73124; CAAS1623.1; -;
 DR EMBL; Z99123; CAB51813.1; -;
 DR FIR; S39722; S39722.
 DR Subtilist; BG10613; spsE.
 DR InterPro; IPR006014; Antifreeze dom.
 DR InterPro; IPR006190; Antifreeze_like.
 DR InterPro; IPR004144; Neut.
 DR Pfam; PF01354; Antifreeze; 1.
 DR Pfam; PF01302; Neut; 1.
 DR PROSITE; PS0844; AFP_LIKE; 1.
 KW Complete proteome.
 FT DOMAIN 305 367 AFP-LIKE.
 SQ SEQUENCE 373 AA; 40889 MW; 3CFBFA6CE0DBCE8 CRC64;
 Query Match 19.6%; Score 370; DB 1; Length 373;
 Best Local Similarity 30.1%; Pred. No. 7.8e-21;
 Matches 107; Conservative 55; Mismatches 167; Indels 26; Gaps 9;
 QY 13 VGGHPCFIITAEIQNHQGLDVAKRMIRMAKCGACAKFKSELEFKNRKALERYPT 72
 DB 11 VGKDAPVFIIAAGINHDGKLQAFALIDAAEAGADAVKFM-----FQADRMVYKQDPL 66
 QY 73 SKHSWGTYGHEK--RHLEFSDHQYRELQVAREVGIFFTASGMDENAVELHNLNPPFF 130
 DB 67 YKTAAGKDVIFSLVQSNMPEAWILPLDYCREKQVIFLSTVDCGSDADLQSTPSAF 126
 QY 131 KVGSGDTNPNPYLEKTAKGRPMVVISGMSQMDTMKVQIVKPL-NPNFCFLOCTSAYP 189
 DB 127 KIASYEINHLPLKYYVARLNRPMTFTAGAEISDVHEAWRTIRAEQNNQTAIMCVAKYP 186
 QY 190 LQEDVNLRVISEYKQLPDPIDIGYSGH-ETGIAISVAALGAKVLERHITLDTWKGS 248
 DB 187 APPEYNLSNIVPMLAAAPPEAVIGFSDHSHPTAPCAAVRLGAKLIEKFTIDKNLPGA 246
 QY 249 DHSASLEPGLAEIVRSV-----LVERALGSPTRKQLLPCEMACNEKLGKSV 295
 DB 247 DHSFALNPDELKEMVDGIRTEALKQITKPVSEKLLGSSYKTTTAEIGEIRNFAYRGI 306
 QY 296 VAKVKIPEGTILTMVLT-VKGE-KYAPPEPDIENLV--GKKVLVTVEEDDTIM 346
 DB 307 FTTAPIQKGEAFSEDNIAVLRLPGKQGLHPR-FFELLTSGVRAVRDIPADTGIV 360

RESULT 4
 ID SMIB MOUSE STANDARD; PRT; 1248 AA.
 AC Q920F6;
 DT 10-OCT-2003 (Rel. 42, Created)
 DT 10-OCT-2003 (Rel. 42, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Structural maintenance of chromosomes 1-like 2 protein (SMC1beta
 DE protein).
 DE protein.
 GN Mus musculus (Mouse).
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A., PARTIAL SEQUENCE, FUNCTION, CHARACTERIZATION, AND
 RP INTERACTION WITH SMC3.
 RC TISSUE-Testis;
 RX MEDLINE=21448993; PubMed=11564881;
 RA Revenkova E., Eijpe M., Heyting C., Gross B., Jessberger R.;
 RT "Novel meiosis-specific isoform of mammalian SMC1.";
 RL Mol. Cell. Biol. 21:6984-6998(2001).
 CC -!- FUNCTION: Meiosis specific component of cohesin complex. The
 CC cohesin complex is required for the cohesion of sister chromatids
 CC after DNA replication. The cohesin complex apparently forms a

FT SITE 44 44 (BY SIMILARITY).
FT SITE 79 79 IMPORTANT FOR ICE-BINDING
FT SITE 84 84 (BY SIMILARITY).
FT SITE 88 88 IMPORTANT FOR ICE-BINDING
FT SITE 114 114 (BY SIMILARITY).
FT STRAND 4 7
FT STRAND 11 11
FT TURN 19 21
FT STRAND 22 25
FT STRAND 33 33
FT HELIX 37 40
FT TURN 41 42
FT STRAND 45 45
FT STRAND 49 49
FT TURN 51 52
FT STRAND 53 54
FT TURN 57 59
FT STRAND 60 60
SQ SEQUENCE 134 AA; 14480 MW; F7F208BF3E2CAA54 CRC64;

Query Match 5.8%; Score 110; DB 1; Length 134;
Best Local Similarity 37.4%; Pred. No. 0.13;
Matches 34; Conservative 12; Mismatches 37; Indels 8; Gaps 2;

Qy 269 VERALGSPTKOLLPCMACNEKLG-----KSVAKVXIPEGITLTMDBLTVKVGEPKAY 322
Db 45 VNRAV--PLGTTLPMDVMKYNEDGTTSPGLSKVANQLPINTALTIVMMKAEVSPKGI 102

Qy 323 PPEDIFNLVKGKVLTVVEEDDTIMEELVDNH 353
Db 103 PSEBISKLVGQVNRAYVLDQTLMPDMVKNY 133

RESULT 6
PARE MYCGE STANDARD; PRT; 633 AA.
AC P47445;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Topoisomerase IV subunit B (EC 5.99.1.1-).
GN PARE OR MG203.
OS Mycoplasma genitalium.
OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.
OX NCBI_TaxID=2097;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=ATCC 33530 / G-37;
RX MEDLINE=96026346; PubMed=7569993;
RA Fraser C.M., Gocayne J.D., White O., Adams M.D., Clayton R.A.,
RA Fleischmann R.D., Bult C.J., Kerlavage A.R., Sutton G., Kelley J.M.,
RA Fritchman J.L., Weidman J.F., Small K.V., Sandusky M., Fuhrmann J.L.,
RA Nguyen D.T., Uterback T.R., Saudek D.M., Phillips C.A., Merrick J.M.,
RA Tomb J.P., Dougherty B.A., Bitt K.F., Hu P.-C., Lueger T.S.,
RA Peterson S.N., Smith H.O., Hutchison C.A. III, Venter J.C.,
RT "The minimal gene complement of Mycoplasma genitalium.";
RN Science 270:397-403(1995).
[2]
RP SEQUENCE OF 406-633 FROM N.A.
RC STRAIN=ATCC 33530 / G-37;
RC Bailey C.C., Younkins R., Huang W.M., Bott K.F.;
RL Submitted (MAY-1995) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: Topoisomerase IV is essential for chromosome
CC segregation. It has relaxation of supercoiled DNA activity.
CC Performs the decatenation events required during the replication
CC of a circular DNA molecule (By similarity).
CC -!- SUBUNIT: Composed of two subunits: parC and parE.

CC -!- SIMILARITY: Belongs to the type II topoisomerase family.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: U39700; AAC71421.1; -;
CC EMBL: U25549; AAC43990.1; -;
CC EIR: D64222; D64222.
CC HSSP: P06982; 1AUJ6.
CC TIGR: MG203; -;
CC InterPro: IPR003594; ATPbind ATPase.
CC InterPro: IPR002288; DNA_gyraseB_C.
CC InterPro: IPR001241; DNA_topoisomII.
CC InterPro: IPR005740; TopoIV_B_Gpos.
CC InterPro: IPR006171; Toprim_dom.
CC Pfam: PF00204; DNA_gyraseB; 1.
CC Pfam: PF00986; DNA_gyraseB; 1.
CC Pfam: PF02518; HATPase_C; 1.
CC Pfam: PF01751; Toprim; 1.
CC PRINTS: PR00418; TP12FAMILY.
CC ProDom: PD149633; DNA_gyraseB_C; 1.
CC SMART: SM00387; HATPase_C; 1.
CC SMART: SM00433; TOP2c; 1.
CC TIGRFAMs: TIGR01058; parE_Gpos; 1.
CC PROSITE: PS00177; TOPOISOMERASE-II; 1.
KW Isomerase; Topoisomerase; ATP-binding; Complete proteome.
SQ SEQUENCE 633 AA; 71569 MW; 249B854169A8CB6 CRC64;

Query Match 5.7%; Score 108.5; DB 1; Length 633;
Best Local Similarity 18.1%; Pred. No. 1.1;
Matches 77; Conservative 63; Mismatches 139; Indels 147; Gaps 18;

Qy 26 GQHQGDLDAKRMIRMAKECG-----ADCAKFOK-----ELEFK 61
Db 137 GQYQSVVQAGKIIQAKKIGDTTSHGTTVSFHADPKVFKAQFDSNIKSRKLSFL 196

Qy 62 ENKALERPYTSKHSWGKTYGEHKLHLEFSDHQRELQRYAEVIGIIFTAGMDMAVEF 121
Db 197 FAKLKL-----TFTDQK-----TNKTTVFSTSL-----VQF 224

Qy 122 LHELVNPFVKVSGDTNNFPYLEKTAKKGR-----PMVSSGMSQMDTMKVYQIVKPL 175
Db 225 LDEINNTVETLG-----QKLLKGEKDGIEVEVVFQFNQSDQETILSFANSIKTF 274

Qy 176 -----NPNFCFL-----QCTSAYPELPQEDVNLRVISEYQKL-----FPDIPGVSG 216
Db 275 EGGSHENGFCIAISDVINSYCRKYNLLKEDKQFQSEIRQGLNAILKVNLPKKNIAFEG 334

Qy 217 HETGIAISVAVALGAKVLERH-----ITLDTWKSGDSHASLEPEGLAEV 263
Db 335 QTSKLSKFSKVKVNVYELVQOQHYFQFLERNNDKLIIDKL-----LNARKIKEQI 385

Qy 264 RSVRLVERALGSPTKO-----LLPCE-----MACNEKLGKSVAVAKVI 301
Db 386 KQRELKKSLSQSKILFLGKLAPCQTKTSKELFIVEGDSAGGTAKMGRDRIFAIL 445

Qy 302 P-EGTILTMDLTVKVGEPKAYPPEDIFNL-----VGKKVLTVVEEDDTIMEELV-----DN 352
Db 446 PLRGKVLNVEKINK---KEAITNEEILTLFCIGTILTNFNKILKYLKGIIMTDADN 502

Qy 353 HGKKIK 358
Db 503 DGAHIQ 508

RESULT 7
ANP2_AUSBR
ID ANP2_AUSBR STANDARD; PRT; 63 AA.

```

AC P12101;
DT 01-OCT-1989 (Rel. 12, Created)
DT 01-OCT-1989 (Rel. 12, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Antifreeze peptide AB2.
OS Austrolycichthys brachycephalus (Antarctic eel pout).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC Acanthomorpha; Acanthopterygii; Percomorpha; Perciformes; Zoarcoidei;
OC Zoarcidae; Austrolycichthys.
OX NCBI_TaxID=8195;
RN [1]
RP SEQUENCE.
RX MEDLINE=89323219; PubMed=2752054;
RA Cheng C.-H.; Devries A.L.;
RT "Structures of antifreeze peptides from the antarctic eel pout,
RT Austrolycichthys brachycephalus."
RL Biochim. Biophys. Acta 997:55-64(1989).
CC -!- FUNCTION: Antifreeze proteins lower the blood freezing point by
CC absorbing ice and inhibiting its growth.
CC -!- SIMILARITY: Belongs to the type-III AFP family.
CC -!- SIMILARITY: Contains 1 AFP-like domain.
DR PIR: S04974; S04974.
DR HSP: P35753; 3RD.
DR InterPro: IPR006014; Antifreeze dom.
DR InterPro: IPR006190; Antifreeze_1like.
DR InterPro: IPR006013; AntifreezeII.
DR Pfam: PF01354; Antifreeze; 1.
DR PRINTS: PR00357; ANTIFREEZEIII.
DR ProDom: PD003258; AntifreezeII; 1.
DR PROSITE: PS00844; AFP_LIKE; 1.
KW Antifreeze protein.
FT DOMAIN 3 62 APP-LIKE.
FT SITE 8 8 IMPORTANT FOR ICE-BINDING
FT SITE 13 13 (BY SIMILARITY).
FT SITE 13 13 IMPORTANT FOR ICE-BINDING
FT SITE 17 17 (BY SIMILARITY).
FT SITE 17 17 IMPORTANT FOR ICE-BINDING
FT SITE 43 43 (BY SIMILARITY).
FT SITE 43 43 IMPORTANT FOR ICE-BINDING
FT SITE 43 43 (BY SIMILARITY).
FT SEQUENCE 63 AA; 7001 MW; A1E7F954598DD061 CRC64;

Query Match 5.7%; Score 108; DB 1; Length 63;
Best Local Similarity 42.6%; Pred. No. 0.07;
Matches 26; Conservative 10; Mismatches 25; Indels 0; Gaps 0;

QY 293 KSVVAKVPEGTILTMVGVKPKYPPEDIFNLGKVLVTVEEDDTIMEELVDN 352
DB 2 KSVVQNQLIPINTALTILVMKAEVSPKGPABEIPRLVGMQVNRVYLDETLMPDWKN 61

QY 353 H 353
DB 62 Y 62

RESULT 8
AROF THEME STANDARD; PRT; 338 AA.
AC Q9WYH8;
DT 15-MAR-2004 (Rel. 43, Created)
DT 15-MAR-2004 (Rel. 43, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Phospho-2-dehydro-3-deoxyheptonate aldolase (EC 2.5.1.54) (Phospho-2-
DE keto-3-deoxyheptonate aldolase) (DAHP synthetase) (3-deoxy-D-arabino-
DE heptulosonate 7-phosphate synthase).
GN AROF OR TM0343.
OS Thermotoga maritima.
OC Bacteria; Thermotogae; Thermotogales; Thermotogaceae; Thermotoga.
OX NCBI_TaxID=2336;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=MSB8 / DSM 3109 / ATCC 43589;

```

```

RX MEDLINE=99287316; PubMed=10360571;
RA Nelson K.E., Clayton R.A., Gill S.R., Gwinn M.L., Dodson R.J.,
RA Haft D.H., Hickey E.K., Peterson J.D., Nelson W.C., Ketchum K.A.,
RA McDonald L., Uterback T.R., Malek J.A., Linher K.D., Garrett M.M.,
RA Stewart A.M., Cotton M.D., Pratt M.S., Phillips C.A., Richardson D.,
RA Heidelberg J., Sutton G.G., Fleischmann K.D., Eisen J.A., White O.,
RA Salzberg S.L., Smith H.O., Venter J.C., Fraser C.M.;
RT "Evidence for lateral gene transfer between Archaea and Bacteria from
RT genome sequence of Thermotoga maritima."
RL Nature 393:323-329(1999).
RN [2]
RP CHARACTERIZATION.
RX STRAIN=MSB8 / DSM 3109 / ATCC 43589;
RX MEDLINE=22753848; PubMed=12743122;
RA Wu J., Howe D.L., Woodard R.W.;
RT "Thermotoga maritima 3-deoxy-D-arabino-heptulosonate 7-phosphate
RT (DAHP) synthase: the ancestral eubacterial DAHP synthase?";
RL J. Biol. Chem. 278:27525-27531(2003).
CC -!- FUNCTION: Catalyzes the condensation of phosphoenolpyruvate (PEP)
CC and D-erythrose-4-phosphate (E4P) giving rise to 3-deoxy-D-
CC arabino-heptulosonate-7-phosphate (DAHP).
CC -!- CATALYTIC ACTIVITY: Phosphoenolpyruvate + D-erythrose 4-phosphate
CC + H(2)O = 3-deoxy-D-erythro-hept-2-ulosonate 7-phosphate +
CC phosphate.
CC -!- COFACTOR: Requires divalent ions.
CC -!- ENZYME REGULATION: Inhibited by L-phenylalanine and L-tyrosine.
CC -!- PATHWAY: Aromatic amino acids biosynthesis; shikimate pathway;
CC first step.
CC -!- SUBUNIT: Homotetramer.
CC -!- MISCELLANEOUS: Optimal pH is 6.3 at 60 degrees Celsius. Maximal
CC activity is at 90 degrees Celsius. Extremely thermostable.
CC -!- SIMILARITY: Belongs to class-I DAHP synthetase family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC -----
CC EMBL; AE001715; AAD35429.1; -.
CC FIR; E72388; E72388.
CC TIGR; TM0343; -.
CC InterPro: IPR006218; DAHP1/KDSA.
CC InterPro: IPR006268; DAHP syn 2.
CC Pfam: PF00793; DAHP_synth_1; 1.
CC TIGRFAMS; TIGR01361; DAHP_synth_Bsub; 1.
CC KW Aromatic amino acid biosynthesis; Transferase; Complete proteome.
CC SQ SEQUENCE 338 AA; 37378 MW; E9634B6704D3DF4D CRC64;

Query Match 5.7%; Score 108; DB 1; Length 338;
Best Local Similarity 22.4%; Pred. No. 0.57;
Matches 54; Conservative 50; Mismatches 91; Indels 46; Gaps 10;

QY 56 SELEFKFNKALRBPYTSKHSWGTQYGEHKRHLFSDHQYRELQRYAEVGVIFFTAGMD 115
DB 118 SELGVKVLGGAYKPRTPSPYSF---QGLGKGLLEY-----LREADKYGMVVTBALG 167
QY 116 EMAVEFELHNLVFFKVGSGDTNNFVLEKTKAGRPDWISSGQSMQMDTKOVQIVKPL 175
DB 168 EDDLPKVAEY-ADIIQIGARNAQNFRLLSKAGSYNKFPVLKRGF--MNTIEFLLSAYEI 224
QY 176 --NPNFCFLQC-----TSAYLPQPEDVNLRLVISEYOKLFPDIPIGYSG- 216
DB 225 ANSGNTKILICERGIRTFEXATRNLTLDISAVPIRKESHLPL-----VDPESHSG 275
QY 217 -HETGIAISVAVALGAK--VLEHITFLDTKMGSDHSASLEPGELAEVRSVLREAL 273
DB 276 RRDILVPLSRAAIAVGAHGIIVEVHPPEKAL--SDGKQLDFELFKELVQEMKKIADAL 333
QY 274 G 274

```



```

Db          334 G 334
RESULT 9
GATB_AQUAE STANDARD; PRT; 478 AA.
AC O66766;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Aspartyl-glutamyl-tRNA(Asn/Gln) amidotransferase subunit B
DE (EC 6.3.5.-) (Asp/Glu-ADT subunit B).
GN GATB OR AQ 461.
OS Aquifex aeolicus.
OC Bacteria; Aquificae; Aquificales; Aquificaceae; Aquifex.
OX NCBI_TaxID=63363;
RN SEQUENCE FROM N.A.
RP STRAIN=VFS;
RC MEDLINE=98196666; PubMed=9537320;
RA Deckert G., Warren P.V., Gaasterland T., Young W.G., Lenox A.L.,
RA Graham D.E., Overbeek R., Snead M.A., Keller M., AuJay M., Huber R.,
RA Feldman R.A., Short J.M., Olson G.J., Swanson R.V.;
RT "The complete genome of the hyperthermophilic bacterium Aquifex
RL Nature 392:353-358(1998).
CC -!- FUNCTION: Allows the formation of correctly charged Asn-tRNA(Asn)
CC or Gln-tRNA(Gln) through the transamidation of misacylated Asp-
CC tRNA(Asn) or Glu-tRNA(Gln) in organisms which lack either or both
CC of asparaginyl-tRNA or glutamyl-tRNA synthetases. The reaction
CC takes place in the presence of glutamine and ATP through an
CC activated phospho-Asp-tRNA(Asn) or phospho-Glu-tRNA(Gln) (By
CC similarity).
CC -!- CATALYTIC ACTIVITY: ATP + L-glutamyl-tRNA(Gln) + L-glutamine = ADP
CC + phosphate + L-glutamyl-tRNA(Gln) + L-glutamate.
CC -!- CATALYTIC ACTIVITY: ATP + L-aspartyl-tRNA(Asn) + L-glutamine = ADP
CC + phosphate + L-asparaginyl-tRNA(Asn) + L-glutamate.
CC -!- SUBUNIT: Heterotrimer of A, B and C subunits (By similarity).
CC -!- SIMILARITY: Belongs to the gatB/gatE family. GatB subfamily.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/annouce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AE000691; AAC06727.1; -.
CC PIR; B70342; B70342.
CC HAMAP; MF_00121; -.
CC InterPro; IPR004413; GatB.
CC InterPro; IPR006107; GatB_cent.
CC InterPro; IPR006075; GatB_N.
CC InterPro; IPR003789; GatB_Yqey.
CC Pfam; PF01162; GatB; 1.
CC Pfam; PF02934; GatB_N; 1.
CC Pfam; PF02637; GatB_Yqey; 1.
CC TIGRFAMs; TIGR00133; GatB; 1.
CC PROSITE; PS01234; GATB; 1.
CC Protein biosynthesis; Ligase; Complete proteome.
CC SEQUENCE 478 AA; 55040 MW; F725AE78944BD79A CRC64;
Query Match 5.6%; Score 106.5; DB 1; Length 478;
Best Local Similarity 19.6%; Pred. No. 1.1;
Matches 62; Conservative 54; Mismatches 100; Indels 101; Gaps 14;
Qy 93 DQYRELQRYAEVGIFFTAGMDKAVVEFLHNLNVPFKVSGD-----TNNFPYL 143
Db 171 EKLRLNRYA---GV---SKADMEKGLQRLCDINVSIRPKGSRFGTRVEIKNVSFRFV 223
Qy 144 EKTAKGRPMVISGMSQMDTMKVQYQIVKPLNPNFCFLQCTSAVPLQ----- 191

```

```

Db          224 QKALE-----YEIERQINNVVEGGEVVQETRTFDP-----QTGKTYPMRTKEBAEDRYRFP 274
Qy 192 -PEDVNLRV---ISEYQKLFDPID-----IGYSGHETGIAISVAAVALGAKVLE 236
Db 275 DFDLVPLKVKKEWIEIKKNMPDPQRFERLIKEYGLSEYEAAGILVNHKEVGDFFEEAV 334
Qy 237 RHITLDK---TW-----KG-SHSHASLSEPGELAEIVRSVRLVERALGSPTKQLL 281
Db 335 RHFKEPKGIVNWLINDLLGLLRDKGISIBESVPKPEHLAEVLKLIK----- 380
Qy 282 PCMACNEKLGKSVVAKVKIPETGITLTDMLTVKVGEPKAYPPEDIFNLVGGKVLVTVEE 341
Db 381 --EKVISTKIGKEVIKEM-----VETGK-----TPSOIVEKGLUKQITDENQ 420
Qy 342 DDTIMEELVDNHRGKKIK 358
Db 421 IKELVKIKFEKPKVE 437
RESULT 10
YD76 MYCPN STANDARD; PRT; 1140 AA.
AC P75405;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DE Hypothetical protein MPN376 (A19_orf1140).
GN MPN376 OR MP460.
OS Mycoplasma pneumoniae.
OC Bacteria; Firmicutes; Mollicutes; Mycoplasmatataceae; Mycoplasma.
OX NCBI_TaxID=2104;
RN SEQUENCE FROM N.A.
RP STRAIN=ATCC 29342 / M129;
RX MEDLINE=97105885; PubMed=8948633;
RA Himmelreich R., Hilbert H., Plagens H., Pirkl B., Li B.-C.,
RA Hermann R.;
RT "Complete sequence analysis of the genome of the bacterium Mycoplasma
RT pneumoniae."
RL Nucleic Acids Res. 24:4420-4449(1996).
CC -!- SIMILARITY: IN THE N-TERMINAL SECTION; STRONG, TO M.PNEUMONIAE
CC MPN375.
CC -!- SIMILARITY: IN THE CENTRAL SECTION; STRONG, TO M.PNEUMONIAE
CC MEN374.
CC -!- SIMILARITY: IN THE C-TERMINAL SECTION; STRONG, TO M.PNEUMONIAE
CC MPN373.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/annouce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AE000045; AAB96108.1; -.
CC PIR; S73786; S73786.
CC KW Hypothetical protein; Transmembrane; Complete proteome.
CC TRANSMEM 8 28 POTENTIAL.
CC TRANSMEM 1098 1118 POTENTIAL.
CC SEQUENCE 1140 AA; 130383 MW; 8FA9406C57DD8886 CRC64;
Query Match 5.4%; Score 102.5; DB 1; Length 1140;
Best Local Similarity 21.3%; Pred. No. 6.7;
Matches 84; Conservative 60; Mismatches 167; Indels 83; Gaps 16;
Qy 22 IAEIGQHQQGDLVAKNMIRMA-----KEGADCAKFOK--SELEFKFNKALERP 71
Db 598 VSAIFQDEGGLRDASPINKVATPKSSKTKIANGDASKIEAKANEISYKYH-----YRQN 653
Qy 72 TSKHSMCKTYGEHKLHLEFSDQYRELQRYAEVGIFFTAGMD--EMAVEFLHNLNVP 129
Db 654 FKQASWDKQNSQTSKIVVQSTDLNDRERFQKIDNNLYKVGISETEIKVNAHKVDA-M 712

```


<hr/>					
Best Local Similarity 22.2%; Pred. No. 8.4;					
Matches 70; Conservative 43; Mismatches 114; Indels 88; Gaps 15;					
<hr/>					
Qy	47	GADCAKFOKSELEF----	KFNKALRPERPYTSKHSGWTKYGBHKLHFSDHOYRELQRYA	102	
Dd	476	GADQTLLQLKQIQSHGFN-----	-SWNQFIHHVAGKVSYDMGDFCERNR	523	
Qy	103	BEVGIFFTASGDMAVEFLHELNVPPFKVGGSDTNPNPYLEKTAKRKGRPMVISGMQSM	162		
Dd	524	D-----VLFD-----MD--LIELMOSSLPTIK-----	SLPENLOAQDKGRPTTAGSKIK--	566	
Qy	163	DTMKQVQIVKPLNPFCFLOCTSAI-----	PLOPEDVNLRVISEYQKLFDPDIP	211	
Dd	567	---KQANDLVSTL-----MCKTPTHYIRCIKNPKTPKPOMEESRVRKHQVEYLGENIR	617		
Qy	212	IGYSGHETGAISVAVALGAKVLERHTIDK-----	TWKSGDHSASLEPGELAEVRSVR	267	
Dd	618	VRRAGY-----AYRRVFQKFQRLORVAILTKATPWVRGDEKGCVL-----	HLLQSVN	663	
Qy	268	LVERALGSPTKQLPCPCMACNEKLGKSVAKVIPEGTILTMDLTVKVGEPKAYPPEDI	327		
Dd	664	M-----DSDF-----QLGRSKVF- IKAPESILFLEEEMRERYKDGYARVIQKW	706		
Qy	328	FNLVGKKVLTVEED	342		
Dd	707	RKFVARKKYVQMRED	721		
<hr/>					
RESULT 13					
BGL1	BACSU	ID	STANDARD;	PRT:	469 AA.
AC	P40740;	O32287;			
DT	01-FEB-1995	(Rel. 31, Created)			
DT	30-MAY-2000	(Rel. 39, Last sequence update)			
DE	10-OCT-2003	(Rel. 42, Last annotation update)			
DE	Beta-glucosidase (EC 3.2.1.21) (Gentiobiase) (Cellobiase) (Beta-D-				
DE	glucoside glucohydrolase) (Amygdalase).				
GN	BGLH OR N17D OR BSU3260.				
OS	Bacillus subtilis.				
OC	Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.				
OX	NCBI_TaxID=1423;				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=168 / Marburg;				
RA	MEDLINE=95189730; PubMed=7883710;				
RX	Le Coq D., Lindner C., Krueger S., Steinmetz M., Stuelke J.;				
RT	"New beta-glucoside (bgl) genes in Bacillus subtilis: the bglp gene				
RT	product has both transport and regulatory functions similar to those				
RT	of BglF, its Escherichia coli homolog.";				
RL	J. Bacteriol. 177:1527-1535(1995).				
RN	[2]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=168 / BGSCIAL;				
RA	MEDLINE=95219088; PubMed=7704263;				
RX	Yoshida K.-I., Sano H., Seki S., Oda M., Fujimura M., Fujita Y.;				
RT	"Cloning and sequencing of a 29 kb region of the Bacillus subtilis				
RT	genome containing the hut and wapA loci."				
RL	Microbiology 141:337-343(1995).				
RN	[3]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=168;				
RA	MEDLINE=98044033; PubMed=9384377;				
RX	Kunst F., Ogawara N., Mosser I., Albertini A.M., Alloni G.,				
RA	Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchart S.,				
RA	Borriass R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,				
RA	Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,				
RA	Choi S.K., Codani J.J., Conerton I.F., Cummings N.J., Daniel R.A.,				
RA	Denicot F., Devine K.M., Dutertoft A., Ehrlich S.D., Emerson P.T.,				
RA	Ertian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,				
RA	Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,				
RA	Ghim S.Y., Glaser P., Goffeau A., Golightly E.J., Grandi G.,				
RA	Guiseppi G., Guy B.J., Haga K., Hatech J., Harwood C.R., Henaut A.,				
RA	Hilbert H., Holtsappel S., Hosono S., Hulio M.F., Itaya M., Jones L.,				
<hr/>					
RA	Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,				
RA	Kobayashi Y., Koetter P., Koningsstein G., Krogh S., Kumano M.,				
RA	Kurita K., Lapides A., Lardinois S., Lauber J., Lazarevic V.,				
RA	Lee S.M., Levine A., Liu H., Masuda S., Maubel C., Medigic C.,				
RA	Medina N., Mellado R.P., Mizuno M., Mostl D., Nakai S., Noback M.,				
RA	Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,				
RA	Parrocan E., Pujic P., Purnelle D., Porwollik S., Prescott A.M.,				
RA	Presecan E., Pujić P., Purnelle D., Rapoport G., Rev M., Reynolds S.,				
RA	Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadaie Y.,				
RA	Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,				
RA	Sekiguchi J., Sekowska A., Seror S.J., Serror P., Shin B.S., Soldo B.,				
RA	Sorokin A., Taconci E., Takagi T., Takahashi H., Takemaru K.,				
RA	Takeuchi M., Tamakoshi A., Tanaka T., Terpstra P., Tognoni A.,				
RA	Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,				
RA	Viari A., Wambutt R., Wedler E., Wedler H., Weitzensegger T.,				
RA	Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,				
RA	Yoshida K., Yoshikawa H.F., Zumstein E., Yoshikawa H., Danchin A.;				
RT	The complete genome sequence of the Gram-positive bacterium Bacillus				
RT	subtilis.";				
RL	Nature 390:249-256(1997).				
CC	- - - - - CATALYTIC ACTIVITY: Hydrolysis of terminal, non-re				

321 -----KSGNLLGVXNPKLSEWGWQIDPKGLRITLNTLYDRYQKP-LFIVENGSL 371

329 NLVGGKVLVTVEEDDTRIMBELVDNH 353
372 GAVDK-----VEBDGTQDDYRINY 391

RESULT 14

MYIE HUMAN
ID MYIE HUMAN STANDARD; PRT; 1109 AA.
AC Q12965;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Myosin Ie (Myosin Ic).
GN MYOIE OR MYOIC.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE OF 1-245 FROM N.A.
RX MEDLINE=95018277; PubMed=7932763;
RA Bement W.M., Wirth J.A., Mooseker M.S.;
RT "Cloning and mRNA expression of human unconventional myosin-IC. A
RT homologue of aneoboid myosin-I with a single IQ motif and an SH3
RT domain.";
RL J. Mol. Biol. 243:356-363(1994).
CC -1- FUNCTION: Myosins are actin-based motor molecules with ATPase
CC activity. Unconventional myosins serve in intracellular movements.
CC Their highly divergent tails are presumed to bind to membranous
CC compartments, which would be moved relative to actin filaments (By
CC similarity).
CC -1- SIMILARITY: BELONGS TO THE MYOSIN HEAVY CHAIN FAMILY.
CC -1- SIMILARITY: Contains 1 IQ domain.
CC -1- SIMILARITY: Contains 1 SH3 domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U14391; AAA62667.1; --
DR PIR; S53601.
DR HSP; P08799; 1MND.
DR Genew; HGNC:7599; MYOIE.
DR MIM; 601479; --
DR GO; GO:0015629; C:actin cytoskeleton; TAS.
DR GO; GO:0003774; F:motor activity; TAS.
DR GO; GO:0008570; F:myosin ATPase activity; TAS.
DR InterPro; IPR000048; IQ region.
DR InterPro; IPR001609; myosin_head.
DR InterPro; IPR001452; SH3.
DR Pfam; PF00612; IQ; 1.
DR Pfam; PF00663; myosin_head; 1.
DR Pfam; PF00018; SH3; 1.
DR PRINTS; PR00193; MYOSINHEAVY.
DR PRINTS; PR00452; SH3DOMAIN.
DR ProDom; PD000355; myosin_head; 1.
DR ProDom; PD000066; SH3; 1.
DR SMART; SM00242; MYSC; 1.
DR SMART; SM00326; SH3; 1.
DR PROSITE; PS50096; IQ; 1.
DR PROSITE; PS50002; SH3; 1.
DR PROSITE; PS50002; SH3; 1.
KW Myosin, ATP-binding; Actin-binding; Calmodulin-binding; SH3 domain;
KW Multigene family.
FT DOMAIN 1 679 MYOSIN HEAD-LIKE.
FT DOMAIN 695 724 IQ.
FT DOMAIN 1052 1109 SH3.
FT DOMAIN 581 591 ACTIN-BINDING (POTENTIAL).
FT

FT NP BIND 112 119 ATP (POTENTIAL).
SQ SEQUENCE 1109 AA; 127040 MW; 0C1C74BAA3705905 CRC64;

Query Match 5.3%; Score 100; DB 1; Length 1109;
Best Local Similarity 22.5%; Pred. NO. 10;
Matches 64; Conservative 41; Mismatches 91; Indels 88; Gaps 15;

QY 47 GADCAKFKSELEF-----KFNKALERPVTSKHSMGKTYGEHKKHLEFSDQYRELQRYA 102
DB 476 GADQTLQLQMQIGSHEFN-----SWNQGFIIHHYAGKVSYDMDFGCERNR 523
QY 103 EEVGIFFTAGMDMAVEFLHNLNVEFFKVGSGDTNNFFLYEKTAKKGRPMVSSGMQSM 162
DB 524 D---VLF---MD--LIELMQSSELPFIK-----SLFPENLQADKKGRPTTAGSKIK-- 566
QY 163 DTMKQVQIVKPLNPNFCLOCTSA-----PLQPEDVNLRVISEYQKLPDIP 211
DB 567 ---KQANDLVSTL-----MKCTPHYIRCIKPNETKKPRDWESRVKHQVEYLGLENIR 617
QY 212 IGVSGHETGIATISVAALGAKVLERHITLTK-----TWKSGDSHSASLEPGELAEVRSVR 267
DB 618 VRRAGY-----AYRIFQKFLQRYAILTKATWPSWQGEKQGVL-----HLQSVN 663
QY 268 LVERALGSPKQLLPCEMACNEKLGKSVVAKVIPEGTILTMDM 311
DB 664 M-----DSDQF-----QLGRSKVP-IKAPESLFLLEEM 690

RESULT 15

GPR BACME
ID GPR BACME STANDARD; PRT; 371 AA.
AC P2321;
DT 01-AUG-1991 (Rel. 19, Created)
DT 01-AUG-1991 (Rel. 19, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Germination protease precursor (EC 3.4.24.78) (Spore protease) (GPR
DE endopeptidase) (Germination proteinase).
GN Bacillus megaterium.
OS Bacillia, Firmicutes; Bacilliales; Bacillaceae; Bacillus.
OC Bacteria, Firmicutes; Bacilliales; Bacillaceae; Bacillus.
OX NCBI_TaxID=1404;
RP SEQUENCE FROM N.A., AND SEQUENCE OF 16-31.
RC STRAIN=ATCC 12872 / QMRDC B1551;
RX MEDLINE=91100296; PubMed=1840582;
RA Sussman M.D., Setlow P.;
RT "Cloning, nucleotide sequence, and regulation of the Bacillus
RT subtilis gpr gene, which codes for the protease that initiates
RT degradation of small, acid-soluble proteins during spore
RT germination.";
RL J. Bacteriol. 173:291-300(1991).
RN [2]
RP CHARACTERIZATION.
RX MEDLINE=98422459; PubMed=9748439;
RA Nessi C., Jedrejas M.J., Setlow P.;
RT "Structure and mechanism of action of the protease that degrades
RT small, acid-soluble spore proteins during germination of spores of
RT Bacillus species.";
RL J. Bacteriol. 180:5077-5084(1998).
RN [3]
RP X-RAY CRYSTALLOGRAPHY (3.0 ANGSTROMS) OF ZYMOCEN P46.
RX MEDLINE=20323395; PubMed=10864493;
RA Ponnuraj K., Rowland S., Nessi C., Setlow P., Jedrejas M.J.;
RT "Crystal structure of a novel germination protease from spores of
RT Bacillus megaterium: structural arrangement and zymogen activation.";
RL J. Mol. Biol. 300:1-10(2000).
CC -1- FUNCTION: Initiates the rapid degradation of small, acid-soluble
CC proteins during spore germination.
CC -1- CATALYTIC ACTIVITY: Endopeptidase action with P4 Glu or Asp, P1
CC preferably Glu > Asp, P1', hydrophobic and P2' Ala.
CC -1- SUBUNIT: Homotetramer.
CC -1- DEVELOPMENTAL STAGE: GPR transcription occurs during sporulation
CC in forespore first by sigma-P and then by sigma-G.

```
CC -!- PTM: Autoproteolytically processed. The inactive tetrameric
CC zymogen termed p46 autoprocesses to a smaller form termed p41,
CC which is active only during spore germination.
CC -!- SIMILARITY: Belongs to peptidase family M63.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M55262; AAA22499.1; -.
DR PIR; A39198; A39198.
DR PDB; 1C8B; 03-MAY-01.
DR MEROPS; M63.001; -.
DR HAMAP; MF_00626; -.
DR InterPro; IPR005080; Peptidase_M63.
DR Pfam; PF03418; Peptidase_U3; 1.
DR ProDom; PD041835; Peptidase_M63; 1.
DR TIGRFAMs; TIGR01441; GPR; 1.
KW Hydrolase; Protease; Zymogen; 3D-structure.
FT PROPEP 1 15
FT CHAIN 16 371 GERMINATION PROTEASE.
SQ SEQUENCE 371 AA; 40626 MW; 40ABE62ECA2F935D CRC64;

Query Match 5.2%; Score 98.5; DB 1; Length 371;
Best Local Similarity 20.9%; Pred. No. 3.4;
Matches 63; Conservative 44; Mismatches 120; Indels 75; Gaps 12;

QY 100 RYA--EEVGIPTASGDMEMAVE-----FLHELNV-----PFKVGSGDTNNFP--- 141
DB 71 RYVTLSEVGIREQDTEKQEAMEEVFAKELNPFIKSLNIPDDASCLVWGLNLSVTPDAL 130
QY 142 -----YLEKTAKGRPW--VISS--GMQSDMTMKQVYQIVKPLNPFC 180
DB 131 GPKAVDNLLITRHLPELOPESVDGFRPVSAIVPGVMGNTGIETSDIIFGVVKVNPDI 190
QY 181 F-LQCTSAVPLQPDVNLRLVISEYQKLPDPIPIGYSGHETG---IAISVAVALGAKVLE 236
DB 191 IAI DALAARS I--ERNVATIQISDSGIHPGSGVGNKRKEISYETLGIPIVIAIGIPTVDA 248
QY 237 RHITLDKTWKS DHDASLEPGELAEVLRSVRLVERALGSPTKQLLPCEMACNEKLGKSVV 296
DB 249 VSITS D-----TIDFTLKHFGREMKEQKPSKSLPSPGNTFGE----- 286
QY 297 AKVKIPEGTILTM DMLTVKVGEPKAYPPEDIENLV-----GKKVLVTVEEDDTIMEEL 349
DB 287 -KKKLTEDDLFNEQRQRTYLGIMIGITLPDEEKRLIHEVLAPIGHNLMTPTKQVDMFIEM 345
QY 350 VD 351
DB 346 AN 347
```

Search completed: September 13, 2004, 14:41:30
Job time : 28 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 13, 2004, 14:38:59 ; Search time 32 Seconds
(without alignments)
579.179 Million cell updates/sec

Title: US-10-759-277-4

Perfect score: 1887

Sequence: 1 MPELELCGRWVGQHPFCF.....BEDDTIMEELVDNHGKKIKS 359

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA:*

1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep:*

2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep:*

3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep:*

4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep:*

5: /cgn2_6/ptodata/2/iaa/6C_COMB.pep:*

6: /cgn2_6/ptodata/2/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB ID	Description
1	1887	100.0	359	4	US-09-516-143A-4
2	336.5	17.8	346	4	US-09-495-406-21
3	113.5	6.0	343	4	US-09-107-532A-6711
4	100	5.3	547	4	US-09-648-004-10
5	93.5	5.0	389	4	US-09-134-000C-5846
6	93	4.9	375	4	US-09-252-991A-24216
7	93	4.9	868	2	US-08-907-166-2
8	92.5	4.9	753	4	US-09-391-340-2
9	92.5	4.9	753	4	US-09-543-681A-5022
10	92	4.9	310	4	US-09-498-520A-12
11	91.5	4.8	576	4	US-09-252-991A-25083
12	91	4.8	336	4	US-09-134-001C-3575
13	90	4.8	1211	4	US-09-134-001C-4820
14	89	4.7	831	4	US-09-269-861A-8
15	88.5	4.7	247	4	US-09-134-000C-5157
16	88.5	4.7	414	1	US-09-337-913-1
17	88.5	4.7	414	2	US-08-750-524-1
18	88	4.7	546	2	US-08-492-027A-8
19	87	4.6	679	3	US-08-913-942-15
20	87	4.6	679	4	US-09-268-347-26
21	87	4.6	1456	4	US-09-134-000C-6427
22	86.5	4.6	402	3	US-09-347-801-16
23	86.5	4.6	563	4	US-08-311-731A-142
24	86	4.6	519	4	US-09-489-039A-11591
25	85.5	4.5	382	1	US-07-768-286B-6
26	85.5	4.5	382	1	US-08-487-823B-3
27	85.5	4.5	382	2	US-08-997-040-3

28	85.5	4.5	382	2	US-09-203-237-3	Sequence 3, Appli
29	85.5	4.5	760	3	US-08-928-941D-29	Sequence 29, Appl
30	85.5	4.5	760	4	US-09-280-590A-29	Sequence 29, Appl
31	85.5	4.5	760	4	US-09-892-398-29	Sequence 29, Appl
32	85	4.5	789	4	US-09-134-000C-4939	Sequence 4939, Ap
33	85	4.5	1201	4	US-09-328-352-5890	Sequence 5890, Ap
34	84.5	4.5	303	3	US-08-928-941D-18	Sequence 18, Appl
35	84.5	4.5	303	4	US-09-280-590A-18	Sequence 18, Appl
36	84.5	4.5	303	4	US-09-892-398-18	Sequence 18, Appl
37	84.5	4.5	323	2	US-09-019-216-3	Sequence 3, Appli
38	84.5	4.5	323	4	US-09-249-241-3	Sequence 3, Appli
39	84.5	4.5	392	1	US-07-768-286B-4	Sequence 4, Appli
40	84.5	4.5	405	1	US-08-121-714-7	Sequence 7, Appli
41	84.5	4.5	405	1	US-08-477-108A-7	Sequence 7, Appli
42	84.5	4.5	405	2	US-08-477-112-7	Sequence 7, Appli
43	84.5	4.5	405	5	PCT-US93-08322-7	Sequence 7, Appli
44	84.5	4.5	415	1	US-07-911-531-19	Sequence 19, Appl
45	84.5	4.5	415	1	US-07-693-636A-19	Sequence 19, Appl

ALIGNMENTS

RESULT 1

US-09-516-143A-4
; Sequence 4, Application US/09516143A
; Patent No. 6333182
; GENERAL INFORMATION:
; APPLICANT: Human Genome Sciences, Inc.
; TITLE OF INVENTION: Human Glycosylation Enzymes
; FILE REFERENCE: PF505PCT
; CURRENT APPLICATION NUMBER: US/09/516,143A
; CURRENT FILING DATE: 2000-03-01
; PRIOR APPLICATION NUMBER: 60/122,409
; PRIOR FILING DATE: 1999-03-02
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
; LENGTH: 359
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-516-143A-4

Query Match 100.0%; Score 1887; DB 4; Length 359;
Best Local Similarity 100.0%; Pred. No. 2.8e-193;
Matches 359; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy	1	MPELELCGRWVGQHPFCFIIAETGQNHQGDLDVAKEMIRMAKCGADCAKFKSELEF	60
Db	1	MPELELCGRWVGQHPFCFIIAETGQNHQGDLDVAKEMIRMAKCGADCAKFKSELEF	60
Qy	61	KFNKALERPVTSKHSMGKTYGEHKRHLEFSDQYRELQRYAEVGIFFTAGMDMAVE	120
Db	61	KFNKALERPVTSKHSMGKTYGEHKRHLEFSDQYRELQRYAEVGIFFTAGMDMAVE	120
Qy	121	FLHELNVPFFKVGSDTNNFPYLEKTAKKGRPMWISSQMSMDTKQYQIVKPLNPNFC	180
Db	121	FLHELNVPFFKVGSDTNNFPYLEKTAKKGRPMWISSQMSMDTKQYQIVKPLNPNFC	180
Qy	181	FLOCTSAYPELOPEDNLRVISEYOKLPDIPIGYSGHETGTAISVAALGAKVLERHIT	240
Db	181	FLOCTSAYPELOPEDNLRVISEYOKLPDIPIGYSGHETGTAISVAALGAKVLERHIT	240
Qy	241	LDTKWGDHSDHASLEPGELAEVLVRSVRLVERALGSPTKQLLPCMACNEKLGKSVAKVK	300
Db	241	LDTKWGDHSDHASLEPGELAEVLVRSVRLVERALGSPTKQLLPCMACNEKLGKSVAKVK	300
Qy	301	IPEGTILTMDLTVKVGEPKAYPPEDIFNLVGKKVLVTVEEDDTIMEELVDNHGKKIKS	359
Db	301	IPEGTILTMDLTVKVGEPKAYPPEDIFNLVGKKVLVTVEEDDTIMEELVDNHGKKIKS	359

RESULT 2

COMPUTER READABLE FORM: CD-ROM ISO9660
MEDIUM TYPE: CD-ROM
COMPUTER: PC


```

; CURRENT APPLICATION NUMBER: US/08/907,166
;
; CURRENT FILING DATE: 1997-08-06
;
; NUMBER OF SEQ ID NOS: 12
;
; SOFTWARE: FastSEQ for Windows Version 3.0
;
; SEQ ID NO 2
;
; LENGTH: 868
;
; TYPE: PRT
;
; ORGANISM: Ammonifex degensii
US-08/907-166-2

```

Query Match	4.9%;	Score 93;	DB 2;	Length 868;				
Best Local Similarity	19.8%;	Pred. No. 1.2;						
Matches	72;	Conservative 55;	Mismatches 102;	Indels 134; Gaps 17;				
QY	51	AKFOKSLSEKFNKALERP	YTSKHSHGKTYGEK	HRLEFSDHOYRELQRYAEVSGIFTT 110				
DB	57	AAPDKSTTP	-----RHAAETV----	KAHRPATPDLRQOFNLKEV----- 95				
QY	111	ASGDMEMAVEFLH	ELNVPFFKVGSGDNNF-	PYLEKTAKGRPMWTS	SSQMSMDTKVQ 168			
DB	96	-----LTA	LNVPVVEKEG	EADDLTGTLVDRAE	KEGMOCLIVTG--DLDAL--- 139			
QY	169	YOIVKPLNPNFCF	LOCTSA	YPLQEB-----	DYNLRVISEYOKL-----PFDIPIGY 214			
DB	140	-QLVSPITTVV	LMKRGISEIAV	NEAEVRRFGVTP	QLPDPFKALGADSNI FGLP--- 195			
QY	215	SGHETGIAISVA	AAVALGAKVL	ERHITLDTKTW	KGSDHSASIEP	GELAEVRSRVLVERALG 274		
DB	196	-----GIGPKTA	-----SR	LQSHQSLEKLE----	SKEFPFAKLRETLE--RHKEBAVL 239			
QY	275	SPTQLLPC	EMACNEKL-----	GKSVAKVKL-----	PEGTILTM DML 312			
DB	240	AKKLALIR	ROVPLEBEEI	IRPWP	GNILATLEVF	SRLEFRTIAKRFLELFP	EARLLSASGL 299	
QY	313	T-----VKV	GEPK-----	-----AYPP-----	EDIFNLV	VGK 333		
DB	300	TPSAVR	VKVERPEBL	ERLGE	BLGRQEP	FAALAYPPV	LRKATSSFLALCLG	CEKVFLEGGP 359
QY	334	KVL 336						
DB	360	EVL 362						

```

RESULT 8
US-09-391-340-2
; Sequence 2, Application US/09391340A
; Patent No. 6492511
; GENERAL INFORMATION:
; APPLICANT: Callen, Walter
; APPLICANT: Mather, Eric
; TITLE OF INVENTION: ISOLATION AND IDENTIFICATION OF NOVEL POLYMERASES
; FILE REFERENCE: 09010/027001
; CURRENT APPLICATION NUMBER: US/09/391,340A
; CURRENT FILING DATE: 1999-09-07
; EARLIER APPLICATION NUMBER: US 08/907,166
; EARLIER FILING DATE: 1997-08-06
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 868
; TYPE: PRT
; ORGANISM: Ammonifex degensii
US-09-391-340-2

```

Query Match	4.9%	Score 93	DB 4	Length 868
Best Local Similarity	19.8%	Fred: No. 1.2		
Matches	72	Conservative	55	Mismatches 102; Indels 134; Gaps 17;
Qv	51	AKFOKSELEPFKNRKALERPVTYSKHSWGKTVGEEHKRHLEFSDOYRELQRYAEVGVGFTT	110	
Db	57	AAFOKSKTTF-----RHALAET---KAHRPATDELRQOENLKEV-----	95	
Qv	111	ASGMDKMAVEELHNLNVPFFKVGSGDTNWF---PYLEKTAAGKRPWVSSGMSQMDTMKQV	168	

```

Db      96  -----L T A L N P V V V E K G F E A D D L I G T L V D R A E K G W Q C L I V T G --D L D A L----- 139
Qy      169  Q Y I V K P L N P N F C F L O C T S A Y P L O P E -----D V N L R V I S Y O K L -----F P D I P I C Y 214
Db      140  -Q L V S P L T T V V L M R K G I S E I A V F N E A E V K R R F G V T P R Q L P D F K A L A G D A S D N I P G L P --- 195
Qy      215  S C H E T G I A I S V A A V A L G A K A V L E R H I T L D K T W K G S D H S A S L P E G E L A E L V R S V R L V E R A L G 274
Db      196  -----G I G P K T A-----S R L L Q S H S L E K L E-----S K E F F P A K L R E T L E--R H K E E A V L 239
Qy      275  S P T K Q L L P C E M A C N E K L-----G K S V V A K V K I-----P E C T I L T M D M L 312
Db      240  A K K A L I R R D V P L E E I I R P W P G N I L A T L V F R G L E P R T L A K R F L E L F P E A R L L S A S G L 299
Qy      313  T-----Y K V G E P K-----A Y P P-----E D I F N L V G K 333
Db      300  T P S A V R V K V E P E L E R I G E E L R G E Q F A A L A Y P P V L R R K A T S S F L A L C I G G E K V F L I E G P 359
Qy      334  K V L 336
Db      360  E V L 362

RESULT 9
US-09-543-681A-5022
; Sequence 5022, Application US/09543681A
; Patent No. 6605709
; GENERAL INFORMATION:
; APPLICANT: GARY BRETON
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEIN
; TITLE OF INVENTION: DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 2709.1002-001
; CURRENT APPLICATION NUMBER: US/09/543,681A
; CURRENT FILING DATE: 2000-04-05
; PRIOR APPLICATION NUMBER: US 60/128,706
; PRIOR FILING DATE: 1999-04-09
; NUMBER OF SEQ ID NOS: 8344
; SEQ ID NO 5022
; LENGTH: 753
; TYPE: PRT
; ORGANISM: Proteus mirabilis
US-09-543-681A-5022

```

	Query Match	4.9%; Score 92.5; DB 4; Length 753;
	Best Local Similarity	20.8%; Pred. No. 1.1;
	Matches	69; Conservative 58; Mismatches 105; Indels 99; Gaps 18;
QY	117	MAVEFLHIANVPPFKVGSGDTNPEY-----LEKTAKKGRPMVISSGWS 161 : : : : : : : :
DB	152	IADFLDLNFPVYYDQKEYTLPIYGEOSWSEKKLTBEELKASEGNKEAVIQLDN 211 : : : : : : : :
QY	162	MDTWKQ-----VYQVKPLPNFCLOQTSAYPLOPEDVNLRVISEY----- 203 : : : : : : : :
DB	212	KDSVKQSCKTKNMIRPFIEEKDPQMVLK---YKLISQDINLLFDSAYLNYPEATVLY 267 : : : : : : : :
QY	204	-----QKLPF-----DIPIGYS-----GHETGIAISVAV-----ALGA-----KVLERH 238 : : : : : : : :
DB	268	QIVKGDKNVPLQENSFLANILYLKKSADLAHHDDLKIIEBLNNSNTLSSNYFRNLLEKY 327 : : : : : : : :
QY	239	I-TLDKTKGSDH-----SASLEP-GELAEVRSVRIVERALG---SPTKQLLPCEMAC 287 : : : : : : : :
DB	328	IDTLLLKYPNSPOAMLALANYLKPNSFYNFKALKLVEKAYNIQSPESKJLLAKLYS 387 : : : : : : : :
QY	288	N-EKLGSVWAKVIPBGTILTDMOLT-----VKV-----GEPKAYPPEDIFNLVGKKV 335 : : : : : : : :
DB	398	NSEGVHQNIRKAVSFLENITNDKLTGKSQRELVKIVDFDGASYLKKEBIVNILRESV 446 : : : : : : : :
QY	336	LV-----TVVED-----DTIMBELVDNH 353 : : : : : : : :
DB	447	IKNKSGFTFNQNVSLAHFYADLLDEEVANN 477 : : : : : : : :


```

RESULT 10
US-09-498-520A-12
; Sequence 12, Application US/09498520A
; Patent No. 6613553
; GENERAL INFORMATION:
; APPLICANT: Rock, Charles O
; APPLICANT: Heath, Richard J
; TITLE OF INVENTION: No. 6613553el Encyl Reductases and Methods of Use Thereof
; FILE REFERENCE: SJ-0022
; CURRENT APPLICATION NUMBER: US/09/498,520A
; CURRENT FILING DATE: 2000-02-04
; NUMBER OF SEQ ID NOS: 62
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 12
; LENGTH: 310
; TYPE: PRT
; ORGANISM: Clostridium acetobutylicum
US-09-498-520A-12

```

Query Match	4.9%;	Score 92;	DB 4;	Length 310;	
Best Local Similarity	20.1%;	Pred. No. 0.29;			
Matches	63;	Conservative 48;	Mismatches 88;	Indels 114; Gaps 15	
QY	88	LFESHDOYRELORVAAE---	VGIFFTASGMDMAVEFLHELHNVPPFKVSGSDTNFF----	140	
Db	49	VEYVRDEIRKAKKLDKPGVNI	MLLSDNAEVA-KMVCEEGKVVTTCAGNPGKVIDMW	107	
QY	141	-----PYLEKTA-----	KKGRPMVISSGMOS-----	MDTMKVQVQIUVKPLNPNFC	180
Db	108	KEHDIKVIPTVASVALARMERC	GVADVAAEGCSGHGVELTTMALVPOVDAIN-----	163	
QY	181	FLQCTSAYPLOPEDVNLREVIS	EYOKLFPDIPIGYSGHETGIAISVAALGA-----	232	
Db	164	-----	IPVIAAGGIGDGRGVA---A	AFALGASGVQVGTGR	194
QY	233	-----	KVLE-RHITLDTKWGSHSASLE	PGELAEHLVSRVLVERALGSP	276
Db	195	FLIAKECTWHQYKNKVLKAKDI	TEVTGRSTGTHPVRVLRNKLA---RKYKLMEXEGASP	251	
QY	277	TKOLLPCEMACNEKLGKSVAAKVI	PGSTGLTMDMLTVKVGEPKAYPPEDI	FNILVCKKVL	336
Db	252	EEM-----	EELGRGALPRA-VREGDV---	DNGSVWAGQ-----IAGLINKE---	288
QY	337	VTVEEDDTIMEEL	349		
Db	289	---ETCDEIVESM	298		

```

RESULT 11
US-09-252-991A-25083
; Sequence 25083, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 25083
; LENGTH: 576
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-25083

```

Query Match 4.8%; Score 91.5; DB 4; Length 576;
Best Local Similarity 19.9%; Pred. No. 0.9;
Matches 78; Conservative 41; Mismatches 146; Indels 12

```

QY 9 PORW---VGGQHPCHII---ABIGNHQGDLD-----VAKRMIRMAKECGADCAKQ 54
Db 138 PTGWPAFVAGAHGAALSAPGSFGTWDLSERRGWPAAVWARRMASFPCLAILAAREQ 197
QY 55 KSELEFKENRKALRPYTSKHSWGKTYGEHKHLEFSDHOYRELQRYAEVGIFFFTASGM 114
Db 198 DAVHE-----ERPGPERRA-----SCHLRYSF-----VLLPRNGF 228
QY 115 DEMAVEFLHELNVPPFKVSGDGTNNPPYLEKTAKKGRPMVISGMSQMDTMKQVYQIVK- 173
Db 229 R-----MKAAGSSRRPRGVSLSVPI-RPVLPARGWAMSVTTDASPLIHW 273
QY 174 ----PLNPNFCLOQTSAY-----PLQPEDVNLRVISEYOKLPDPIPIGYS 215
Db 274 SKAMPESLQLMYFACVAEYSGTQAAEVLHVQSOPSLSRQIQALEVHIGALFVRGPAVGS 333
QY 216 GHETGIAISVAVAL--GAKVLERHI-----TLDKTWG----- 247
Db 334 LTEVGRVVHCRAVKIIABAKALTRDIGDDRDLHGFSRYLDPCWSELARQLKEARVC 393
QY 248 SDHSASLEPG-ELAEVLRSVRLVERALSGPTKQLLPCEM-----ACNEKLGSVV 296
Db 394 SAVSPSWAPSLDLLSLVRESSLDAAIIGAPPKVREYELMEIGSYEHVIALSETLYRRLA 453
QY 297 AKVKIPEGTILTMOMLTVKVGEPKAYPPEDIF 328
Db 454 ----PWSLNDL-----IGEPLLVPKDSLF 474

```

```

RESULT 12
US-09-134-001C-3575
; Sequence 3575, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 3575
; LENGTH: 336
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-3575

```

	Query Match	4.8%	Score 91;	DB 4;	Length 336;
	Best Local Similarity	21.7%;	Pred. No. 0.43;		
	Matches	52;	Conservative	32;	Mismatches 64;
					Indels 92;
					Gaps 10;
QY	141	PYLEKTAQGRPMV	ISSGMQMDTKQVQIVKPLNFCFLOCTSA	YAPLPQDPVNL	--- 197
Db	27	PHLE---	NGKVLIGA-----	TTSNPYHAINPAIRSRA--	QIFELYPLQDDIRALD 74
QY	198	-----	RVISEYQKLFDPDIPGYSGHETGIAISVAVALGAKVL	-----	ERHITLD 242
Db	75	RAINDKERGLSTYHP	IVDEDAIEYFTSQSGDVRSALNALEAVLSAHI	GEENERHITLD	134
QY	243	-----	-----	KTWKSGDHSASL-----	EPCELAELVRSVR 267
Db	135	DAKDCLQKAFYSDKGD	MHYDVMSAFQKIRSGDVNAAHYLARIEAGDLPTI	VRLL 194	
QY	268	LV---	FRALGSPYKQLLPCEMACNEKLGKSVAKVI	PEGTILTMOMLTKVKGEPKAYPP	324
Db	195	VISYEDVGLASP-----	-----	NAGORTLAAIQSAE-----	RLGFPPARIP 230


```
QY 137 TNNF-----PYLEKTA-----KGRBPVVISSQMOMDMKQVYQIVKPLNENPCF 181
Db 150 NQNFLLFIILSHSGEEKYLLAKETALLAKEKKHPHIIISFVGAKN-STLGRDLADLVFSTDYSYSPF 208
QY 182 LQCTSAIPLQPEDVNLRLVISEYQKLFDPDIPIGYSGHETGIAIS 224
Db 209 ----STSVAPQPMFFGQTLITFEALI-----CAYLNHEDSIPIS 243
```

Search completed: September 13, 2004, 14:45:00
Job time : 34 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 13, 2004, 14:43:50 ; Search time 130 Seconds
(without alignments)
885.597 Million cell updates/sec

Title: US-10-759-277-4
Perfect score: 1887
Sequence: 1 MPLELELCGRWVGQHPCF.....BEDDTIMEELVDNHGKKIKS 359

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1335176 seqs, 320689617 residues

Total number of hits satisfying chosen parameters: 1335176

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database : Published Applications AA:**

- 1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pcp.*
- 2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pcp.*
- 3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pcp.*
- 4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pcp.*
- 5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pcp.*
- 6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pcp.*
- 7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pcp.*
- 8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pcp.*
- 9: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pcp.*
- 10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pcp.*
- 11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pcp.*
- 12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pcp.*
- 13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pcp.*
- 14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pcp.*
- 15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pcp.*
- 16: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pcp.*
- 17: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pcp.*
- 18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pcp.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1887	100.0	359	9	US-09-984-205-4
2	1887	100.0	359	9	US-09-930-440B-6
3	1887	100.0	359	16	US-10-759-277-4
4	1883	99.8	359	15	US-10-264-237-2597
5	544.5	28.9	338	9	US-09-767-041-24
6	524	27.8	311	14	US-10-156-761-10910
7	510.5	27.1	346	9	US-09-930-440B-8
8	381	20.2	123	14	US-10-106-698-6926
9	336.5	17.8	346	9	US-09-816-028A-35
10	336.5	17.8	346	14	US-10-303-161-35
11	336.5	17.8	346	14	US-10-303-118-35
12	336.5	17.8	346	14	US-10-303-128-35
13	336.5	17.8	346	14	US-10-303-134-35
14	336.5	17.8	346	14	US-10-303-162-35
15	303	16.1	340	10	US-09-882-227-490

16	287	15.2	340	12	US-10-335-977-7626	Sequence 7626, Ap
17	283.5	15.0	314	12	US-10-335-977-7625	Sequence 7625, Ap
18	122.5	6.5	205	12	US-10-321-396C-4	Sequence 4, Appli
19	119	6.3	352	15	US-10-369-493-18790	Sequence 18790, A
20	117.5	6.2	275	12	US-10-321-396C-2	Sequence 2, Appli
21	116	6.1	352	15	US-10-369-493-20092	Sequence 20092, A
22	113.5	6.0	333	12	US-10-282-122A-57496	Sequence 57496, A
23	113	6.0	285	15	US-10-321-396C-6	Sequence 18119, A
24	110	5.8	135	12	US-10-369-493-21568	Sequence 21568, A
25	110	5.8	265	15	US-10-369-493-3746	Sequence 9746, Ap
26	109	5.8	337	15	US-10-369-493-20181	Sequence 20181, A
27	108.5	5.7	335	15	US-10-369-493-2631	Sequence 2631, Ap
28	108.5	5.7	351	15	US-10-369-493-63506	Sequence 63506, A
29	108.5	5.7	633	12	US-10-282-122A-53304	Sequence 53304, A
30	108	5.7	338	15	US-10-369-493-2930	Sequence 2930, Ap
31	107.5	5.7	337	12	US-10-282-122A-51516	Sequence 51516, A
32	107	5.7	276	15	US-10-369-493-22853	Sequence 22853, A
33	107	5.7	330	12	US-10-282-122A-67597	Sequence 67597, A
34	103.5	5.5	329	15	US-10-369-493-13704	Sequence 13704, A
35	101.5	5.4	335	15	US-10-369-493-8945	Sequence 8945, Ap
36	101.5	5.4	337	12	US-10-282-122A-53304	Sequence 53304, A
37	101.5	5.4	341	12	US-10-282-122A-57228	Sequence 57228, A
38	101	5.4	1175	15	US-10-369-493-6504	Sequence 6504, Ap
39	100	5.3	547	14	US-10-272-419-10	Sequence 10, Appl
40	100	5.3	547	14	US-10-230-026-20	Sequence 20, Appl
41	100	5.3	1104	16	US-10-322-281-778	Sequence 778, App
42	100	5.3	1108	12	US-10-087-192-1206	Sequence 1206, Ap
43	100	5.3	1189	12	US-10-282-122A-70920	Sequence 70920, A
44	99.5	5.3	702	16	US-10-437-963-169700	Sequence 169700,
45	98	5.2	715	16	US-10-437-963-120663	Sequence 120663,

ALIGNMENTS

RESULT 1

US-09-984-205-4
; Sequence 4, Application US/09984205
; Patent No. US20020137175A1
; GENERAL INFORMATION:
; APPLICANT: Coleman, Timothy A. et al.
; TITLE OF INVENTION: Human Glycosylation Enzymes
; FILE REFERENCE: PF50SD1
; CURRENT APPLICATION NUMBER: US/09/984,205
; CURRENT FILING DATE: 2001-10-29
; PRIOR APPLICATION NUMBER: PCT/US00/05325
; PRIOR FILING DATE: 2000-03-01
; PRIOR APPLICATION NUMBER: 09/516,143
; PRIOR FILING DATE: 2000-03-01
; PRIOR APPLICATION NUMBER: 60/122,409
; PRIOR FILING DATE: 1999-03-02
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
; LENGTH: 359
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-984-205-4

Query Match 100.0%; Score 1887; DB 9; Length 359;
Best Local Similarity 100.0%; Pred. No. 1.7e-182;
Matches 359; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	MPLELELCGRWVGQHPCFIIAEIGNQHGDLDVAKRMIRMAKEGADCAKFKSELEF	60
Db	1	MPLELELCGRWVGQHPCFIIAEIGNQHGDLDVAKRMIRMAKEGADCAKFKSELEF	60
QY	61	KFNKALERPYSKHSWGKTYGEHKLHLEFSDHQYRELQRYAEVGIFFTASGMDMAVE	120
Db	61	KFNKALERPYSKHSWGKTYGEHKLHLEFSDHQYRELQRYAEVGIFFTASGMDMAVE	120
QY	121	FLHNLNVPFKVSGSDTNFPYLEKTKAGRPVMISSGMSMDTMKVQIVKPLNPNFC	180

Db 121 FLHNLVPPFKVGGSDTNNPYLEKTAKKGRPMVSSGMSMDTMKQVYQIVKPLNPNFC 180
 QY 181 FLQCTSAYPELOPEDVNLRVISEYQKLPDPIPIGYSGHETGTAISVAALGAKVLERHIT 240
 Db 181 FLQCTSAYPELOPEDVNLRVISEYQKLPDPIPIGYSGHETGTAISVAALGAKVLERHIT 240
 QY 241 LDKTWKSDHSASLEPGEALVRSVRLVERALGSPTKQLLPCMACNEKLGKSVVAKVK 300
 Db 241 LDKTWKSDHSASLEPGEALVRSVRLVERALGSPTKQLLPCMACNEKLGKSVVAKVK 300
 QY 301 IPEGTILTMDLTVKVGEPKAYPPEDIFNLVGKVLVTVEEDDTIMEELVDNHGKKIKS 359
 Db 301 IPEGTILTMDLTVKVGEPKAYPPEDIFNLVGKVLVTVEEDDTIMEELVDNHGKKIKS 359

RESULT 2
 US-09-930-440B-6
 ; Sequence 6, Application US/09930440B
 ; Patent No. US20020142386A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Betenbaugh et al.
 ; TITLE OF INVENTION: Engineering Intracellular Sialylation Pathways
 ; FILE REFERENCE: PF509P2
 ; CURRENT APPLICATION NUMBER: US/09/930,440B
 ; PRIOR FILING DATE: 2001-08-16
 ; PRIOR APPLICATION NUMBER: 60/227,579
 ; PRIOR FILING DATE: 2000-08-25
 ; PRIOR APPLICATION NUMBER: 09/516,793
 ; PRIOR FILING DATE: 2000-03-01
 ; PRIOR APPLICATION NUMBER: 60/169,624
 ; PRIOR FILING DATE: 1999-12-08
 ; PRIOR APPLICATION NUMBER: 60/122,582
 ; PRIOR FILING DATE: 1999-03-02
 ; NUMBER OF SEQ ID NOS: 8
 ; SOFTWARE: PatentIn Ver. 2.1
 ; SEQ ID NO 6
 ; TYPE: PRT
 ; LENGTH: 359
 ; ORGANISM: Homo sapiens
 ; ORGANISM: Homo sapiens
 US-09-930-440B-6

Query Match 100.0%; Score 1887; DB 16; Length 359;
 Best Local Similarity 100.0%; Pred. No. 1.7e-182;
 Matches 359; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MPELELCPRWVGQHPCHFIIEIGQNHQGLDVAKRMRMAKEGADCAKFKQKSELEF 60
 Db 1 MPELELCPRWVGQHPCHFIIEIGQNHQGLDVAKRMRMAKEGADCAKFKQKSELEF 60
 QY 61 KFNKRALERTYTSKHSWKTYGHEKHLFESHQYRELQRYAEVGIFFFTASGMDMAVE 120
 Db 61 KFNKRALERTYTSKHSWKTYGHEKHLFESHQYRELQRYAEVGIFFFTASGMDMAVE 120
 QY 121 FLHNLVPPFKVGGSDTNNPYLEKTAKKGRPMVSSGMSMDTMKQVYQIVKPLNPNFC 180
 Db 121 FLHNLVPPFKVGGSDTNNPYLEKTAKKGRPMVSSGMSMDTMKQVYQIVKPLNPNFC 180
 QY 181 FLQCTSAYPELOPEDVNLRVISEYQKLPDPIPIGYSGHETGTAISVAALGAKVLERHIT 240
 Db 181 FLQCTSAYPELOPEDVNLRVISEYQKLPDPIPIGYSGHETGTAISVAALGAKVLERHIT 240
 QY 241 LDKTWKSDHSASLEPGEALVRSVRLVERALGSPTKQLLPCMACNEKLGKSVVAKVK 300
 Db 241 LDKTWKSDHSASLEPGEALVRSVRLVERALGSPTKQLLPCMACNEKLGKSVVAKVK 300
 QY 301 IPEGTILTMDLTVKVGEPKAYPPEDIFNLVGKVLVTVEEDDTIMEELVDNHGKKIKS 359
 Db 301 IPEGTILTMDLTVKVGEPKAYPPEDIFNLVGKVLVTVEEDDTIMEELVDNHGKKIKS 359

RESULT 4
 US-10-264-237-2597
 ; Sequence 2597, Application US/10264237
 ; Publication No. US20040009491A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Birse et al.
 ; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
 ; FILE REFERENCE: PA131P1
 ; CURRENT APPLICATION NUMBER: US/10/264,237
 ; CURRENT FILING DATE: 2002-10-04
 ; PRIOR APPLICATION NUMBER: PCT/US01/16450
 ; PRIOR FILING DATE: 2001-05-18
 ; PRIOR APPLICATION NUMBER: US 60/205,515
 ; PRIOR FILING DATE: 2000-05-19
 ; NUMBER OF SEQ ID NOS: 2876
 ; SOFTWARE: PatentIn Ver. 3.1
 ; SEQ ID NO 2597
 ; LENGTH: 359
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 ; ORGANISM: Homo sapiens
 US-10-264-237-2597

RESULT 3
 US-10-759-277-4
 ; Sequence 4, Application US/10759277

```
Query Match
Best Local Similarity 99.8%; Score 1883; DB 15; Length 359;
Matches 358; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MPLELELCGRVWGQHQHPCFIIAIEIGQNHQGLDVAKRMIRMAKCEGADCAKFKSELEF 60
DB 1 MPLELELCGRVWGQHQHPCFIIAIEIGQNHQGLDVAKRMIRMAKCEGADCAKFKSELEF 60

QY 61 KPNRKALERPYSKHSWGKTYGHEKRLHLEFSDQYRELQRYAEVGIPTFASGMDMAVE 120
DB 61 KPNRKALERPYSKHSWGKTYGHEKRLHLEFSDQYRELQRYAEVGIPTFASGMDMAVE 120

QY 121 FLHNLVVFVKVGGSDTNNPYLEKTAKKGRPMVLSGMSQMDTMKVYQIVKPLNPNFC 180
DB 121 FLHNLVVFVKVGGSDTNNPYLEKTAKKGRPMVLSGMSQMDTMKVYQIVKPLNPNFC 180

QY 181 FLOCTSAVPLQPEDVNLRVISEYQKLFDPDIPIGYSGHETGTAISVAVALGAKVLERHIT 240
DB 181 FLOCTSAVPLQPEDVNLRVISEYQKLFDPDIPIGYSGHETGTAISVAVALGAKVLERHIT 240

QY 241 LDKTWKSDHSALEPGBELVRSVRLVERALGSPTKQLLPCEMACNEKLGKSVVAKVK 300
DB 241 LDKTWKSDHSALEPGBELVRSVRLVERALGSPTKQLLPCEMACNEKLGKSVVAKVK 300

QY 301 IPEGTILTMDLTVKVGEPKAYPPEDIENLYGKKVLVTVEEDDTIMEELVDNHHGKIKS 359
DB 301 IPEGTILTMDLTVKVGEPKAYPPEDIENLYGKKVLVTVEEDDTIMEELVDNHHGKIKS 359

RESULT 5
US-09-767-041-24
; Sequence 24, Application US/09767041
; Patent No. US20020055168A1
; GENERAL INFORMATION:
; APPLICANT: Smith, Hilda
; TITLE OF INVENTION: STREPTOCOCCUS SUIS VACCINES AND DIAGNOSTIC TESTS
; FILE REFERENCE: 2183-4726
; CURRENT APPLICATION NUMBER: US/09/767,041
; CURRENT FILING DATE: 2001-01-22
; PRIOR APPLICATION NUMBER: PCT/NL99/00460
; PRIOR FILING DATE: 1999-07-19
; PRIOR APPLICATION NUMBER: EP98202465.5
; PRIOR FILING DATE: 1998-07-22
; PRIOR APPLICATION NUMBER: EP98202467.1
; PRIOR FILING DATE: 1998-07-22
; NUMBER OF SEQ ID NOS: 53
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 24
; LENGTH: 338
; TYPE: PRT
; ORGANISM: Streptococcus suis
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: CPS2p
US-09-767-041-24

Query Match
Best Local Similarity 28.9%; Score 544.5; DB 9; Length 338;
Matches 123; Conservative 73; Mismatches 129; Indels 17; Gaps 8;

QY 20 FIIAIEIGQNHQGLDVAKRMIRMAKCEGADCAKFKSELEFKNRKALERPYSKHSWGK 79
DB 3 YIIAIEIGCNHNGDVLHARKMVEVAVDCGVDAKFKTEKADLLISKYAPKAEY-QKITGE 61

QY 80 TYG--EHRKRLHLEFSDQYRELQRYAEVGIPTFASGMDMAVEFLHNLVVFVKVGGSDT 137
DB 62 SDSQLEWTRRLLELSFEYLDLRDCLKGVVDVFTPEDESLDFLSTDMVPIKPSGEI 121

QY 138 NNFPYLEKTAKKGRPMVLSGMSQMDTMKVYQIVKPLNPN----FCFLOCTSAVPLQPE 193
DB 122 TNLPLYLEKIGRQAKKVLSTGMVMD--EIHQAVKILQENGTTDISILHCTTETPTYP 178
```

```
QY 194 DVNLRVISEYQKLFDPDIPIGYSGHETGTAISVAVALGAKVLERHITLDKTKWGS DSHS 253
DB 179 ALNVLNVLTLTKKEFPNLITGYSDHSGSEVPITAAAMGAELTEKFTLDNEMEGPDHKAS 238

QY 254 LEFGLELAEVRSVRLVERALGSPTKQLLPCEMACNEKL--GKSUVAKVKIPGTTILTM 311
DB 239 ATPDLAALVKGVRIVEQSLGKFEKE--PEEVEVRNKIVAEKSIYAKKALAKGEVFT 296

QY 312 LTVKVGEP-KAYPPEDIENLYGKKVLVTVEEDDTIMEELVDN 352
DB 297 ITVK--RPGNGISPMWYKVLQGVSEQDPEEDQNICHSAFEN 336

RESULT 6
US-10-156-761-10910
; Sequence 10910, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 10910
; LENGTH: 311
; TYPE: PRT
; ORGANISM: Streptomyces avermitilis
US-10-156-761-10910

Query Match
Best Local Similarity 27.8%; Score 524; DB 14; Length 311;
Matches 122; Conservative 53; Mismatches 115; Indels 24; Gaps 7;

QY 14 GGOHPCFIIAIEIGQNHQGLDVAKRMIRMAKCEGADCAKFKSELEFKNRKA--LERPY 71
DB 15 GPGHPVYVVGIEGINHNGELENAFKLIDAAEAGCDVAKFKQKRTPEICTPRDQWDIER-- 72

QY 72 TSKHSWGK-TYGEKRLHLEFSDQYRELQRYAEVGIPTFASGMDMAVEFLHNLVDPFF 130
DB 73 --DTPWGRMTYIDYRHRVFEGEDEYRQIDEYAKSKNIDWFASPWDTEAVAFLKFPDIPAH 130

QY 131 KVGSGDTNNPYLEKTAKKGRPMVLSGMSQMDTMKVYQIVKPL-NPNFCFLOCTSAYP 189
DB 131 KVASASITDDELLRALRGTRTVILSTG---MSTPKQIRHAEVILGSDNILLCHATSATYP 187

QY 190 LQPEDVNLRVISEYQKLFDPDIPIGYSGHETGTAISVAVALGAKVLERHITLDKTKWGS 249
DB 188 AKABELNLRVINTLQAEPVNPVIGYSGHETGLQTTLAALVALGATFVERHITLDRAMWGS 247

QY 250 HSASLEPGELEAELVRSVRLVERALGSPTKQLLPCEMACNEKLGK--SVVAKVKIPEGTIL 307
DB 248 QAAASVEPQGLTRVDRIRTIEASLGDGVGVKVESELGPMKKLRRVPGVAAEAI----- 301

QY 308 TMDMLTVKVGEPKA 321
DB 302 -----AAAGAPVA 310

RESULT 7
US-09-930-4408-8
; Sequence 8, Application US/099304408
; Patent No. US20020142386A1
```

```

; GENERAL INFORMATION:
; APPLICANT: Betenbaugh et al.
; TITLE OF INVENTION: Engineering Intracellular Sialylation Pathways
; FILE REFERENCE: PF509P2
; CURRENT APPLICATION NUMBER: US/09/930,440B
; PRIOR FILING DATE: 2001-08-16
; PRIOR APPLICATION NUMBER: 60/227,579
; PRIOR FILING DATE: 2000-08-25
; PRIOR APPLICATION NUMBER: 09/516,793
; PRIOR FILING DATE: 2000-03-01
; PRIOR APPLICATION NUMBER: 60/169,624
; PRIOR FILING DATE: 1999-12-08
; PRIOR APPLICATION NUMBER: 60/122,582
; PRIOR FILING DATE: 1999-03-02
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 8
; LENGTH: 346
; TYPE: PRT
; ORGANISM: Escherichia coli
US-09-930-440B-8

Query Match      27.1%; Score 510.5; DB 9; Length 346;
Best Local Similarity 36.3%; Pred. No. 9.4e-43;
Matches 123; Conservative 61; Mismatches 140; Indels 15; Gaps 6;

QY 20 FIIAEGQHQDLDVAKRMIRMAKECGADCAKFKQSELEFKENKALRPPYTSKHSWG- 78
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 5 YIVAEIGCNHGSVDIAREMILKAEAGVNAVKFQTFKADKLISAIAPKAEYQIKNTGEL 64
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 79 KTYGEKRLHLEFSDQYRELQRYAEVGIFFFTASGMDENAVEFLHNLNVPFFKVGSGDNT 138
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 65 ESQLEMTKLEMKYDYLHLMYAVSLNDVSTPDESDIDFLASKOKIKWIPSGELL 124
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 139 NFPYLEKTAKGRP---MVISGQMSMDTMKQYQIV---KPLNPNFCFLOQTSAYPLOP 192
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 125 NLPYLEKIAKLPIPKKIIISTGMATIDEIKQSVSIFINNKPVGNIITILHCNTEYPTPF 184
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 193 EDVNLRVISEYQKLPDIPIGYSGHETGTAISVAALGAKVLERHITLDTKWKSGDHA 252
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 185 EDVNLNAINDLKHFFKNIGFSDHSGFYAAAPVPGITFIEKHTLTKDKNSGPDHLA 244
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 253 SLEPGELAEVRVRIVERALGSPTKQLLPCENACNEKLGKSVVAKVKIPEGTILTMML 312
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 245 STEPDELKHLCTGRCVCKSLGNSKVVTSASERKNIVARKSIIAKTEIKKGVTSEKNI 304
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 313 TVKVGEP-KAYPPEDIFNLVGKVLVTVBEDDTIMEELV 350
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 305 TTK--RPGNGISPMEWNLGK-----IAEQDFIPDELI 336
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

RESULT 8
US-10-106-698-6926
; Sequence 6926, Application US/10106698
; Publication No. US20030109690A1
; GENERAL INFORMATION:
; APPLICANT: Ruben et al.
; TITLE OF INVENTION: Colon and Colon Cancer Associated Polynucleotides and Polypeptides
; FILE REFERENCE: PA005P1
; CURRENT APPLICATION NUMBER: US/10/106,698
; CURRENT FILING DATE: 2002-03-27
; PRIOR APPLICATION NUMBER: PCT/US00/26524
; PRIOR FILING DATE: 2000-09-28
; PRIOR APPLICATION NUMBER: US 60/157,137
; PRIOR FILING DATE: 1999-09-29
; PRIOR APPLICATION NUMBER: US 60/163,280
; PRIOR FILING DATE: 1999-11-03
; NUMBER OF SEQ ID NOS: 8564
; SOFTWARE: PatentIn Ver. 3.0
; SEQ ID NO 6926
; LENGTH: 123
; TYPE: PRT
; ORGANISM: Homo sapiens

```

```

; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (83)-
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: MISC FEATURE
; LOCATION: (87)-
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: MISC FEATURE
; LOCATION: (111)-
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: MISC FEATURE
; LOCATION: (112)-
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: MISC FEATURE
; LOCATION: (121)-
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-10-106-698-6926

Query Match      20.2%; Score 381; DB 14; Length 123;
Best Local Similarity 93.7%; Pred. No. 2.6e-30;
Matches 74; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 183 QCTSAYPLOPQEDVNLRVISEYQKLPDIPIGYSGHETGTAISVAALGAKVLERHITLD 242
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 24 KCTSAYPLOPQEDVNLRVISEYQKLPDIPIGYSGHETGTAISVAALGAKVLERHITLX 83
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 243 KTWKSGDHSASLEPGELAE 261
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 84 KTWKSGDHSASLEPGELGE 102
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

RESULT 9
US-09-816-028A-35
; Sequence 35, Application US/09816028A
; Patent No. US20020042369A1
; GENERAL INFORMATION:
; APPLICANT: Gilbert, Michel
; APPLICANT: Wakarchuk, Warren W.
; APPLICANT: National Research Council of Canada
; TITLE OF INVENTION: Campylobacter Glycosyltransferases for Biosynthesis of
; TITLE OF INVENTION: Gangliosides and Ganglioside Mimics
; FILE REFERENCE: 019633-000111US
; CURRENT APPLICATION NUMBER: US/09/816,028A
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: US 60/118,213
; PRIOR FILING DATE: 1999-02-01
; PRIOR APPLICATION NUMBER: US 09/495,406
; PRIOR FILING DATE: 2000-01-31
; NUMBER OF SEQ ID NOS: 49
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 35
; LENGTH: 346
; TYPE: PRT
; ORGANISM: Campylobacter jejuni
; FEATURE:
; OTHER INFORMATION: sialic acid synthase from C. jejuni OH4384 (ORF 8a
; OTHER INFORMATION: of lipooligosaccharide (LOS) biosynthesis locus)
US-09-816-028A-35

Query Match      17.8%; Score 336.5; DB 9; Length 346;
Best Local Similarity 29.1%; Pred. No. 4.4e-25;
Matches 96; Conservative 63; Mismatches 164; Indels 7; Gaps 4;

QY 21 IIAEIGNHOGDLDVAKRMIRMAKECGADCAKFKQSELEFKENKALRPPYTSKHSWGK 79
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 19 VPEIGININGSLELAKIMVDAAPSTGAKIHKQTHIVEDEMSKAAKKVTPGNAKIS--- 75
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 80 TYGEKRLHLEFSDQYRELQRYAEVGIFFFTASGMDENAVEFLHNLNVPFFKVGSGDTNN 139
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 76 -IYEINQKCALDYKDELALKEYTEKLGVLVLTSPFRAGANRLDEMGVSFAFKIGSGECNN 134
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 140 FVYLEKTAKGRPMVLTSSGQMSMDTMKQYQIVKVLNPNFCFLOQTSAYPLOPQEDVNLRV 199
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```


Wed Sep 15 10:55:31 2004

us-10-759-277-4.rapb

Page 5

```

Db 135 YPLIKHIAAFKPMIVSTGMNSIESIKPTVKILLDNEIPFVLMHTTNLYPTPHNLVRLNA 194
Qy 200 ISEYQKLFDPDIPGYSGHETGIAISVAALGAKVLERHITLDTKWGSDHSALEPGEL 259
Db 195 MLELKKEF-SCMVGLSDHTTDLNACLGAVAGACVLERHFTDSMHRSGPDIVCSMDTQAL 253
Qy 260 AEL-VRSVRLVERALGSPYKQLLPCEMACNEKLGKSVVAKVPIPEGTILTMWMLTVKVG 318
Db 254 KELLIOSEQMAIMRGNNESKAAKQEQVTIDFASFVSVSIKDIKKGEVLSMDNIWVKRPG 313
Qy 319 PKAYPPEDIENLVGKKVLTVEEDDTIMEE 348
Db 314 LGGISAAEFENILGKKALRDIENDTQLSYE 343

RESULT 10
US-10-303-161-35
; Sequence 35, Application US/10303161
; Publication No. US20030148459A1
; GENERAL INFORMATION:
; APPLICANT: Gilbert, Michel
; APPLICANT: Wakarchuk, Warren W.
; APPLICANT: National Research Council of Canada
; TITLE OF INVENTION: Campylobacter Glycosyltransferases for Biosynthesis of
; TITLE OF INVENTION: Gangliosides and Ganglioside Mimics
; FILE REFERENCE: 019633-000111US
; CURRENT APPLICATION NUMBER: US/10/303,161
; CURRENT FILING DATE: 2002-11-21
; PRIOR APPLICATION NUMBER: US/09/816,028
; PRIOR FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: US 60/118,213
; PRIOR FILING DATE: 1999-02-01
; PRIOR APPLICATION NUMBER: US 09/495,406
; PRIOR FILING DATE: 2000-01-31
; NUMBER OF SEQ ID NOS: 49
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 35
; LENGTH: 346
; TYPE: PRT
; ORGANISM: Campylobacter jejuni
; FEATURE:
; OTHER INFORMATION: sialic acid synthase from C. jejuni OH4384 (ORF 8a
; OTHER INFORMATION: of lipooligosaccharide (LOS) biosynthesis locus)
US-10-303-161-35

Query Match 17.8%; Score 336.5; DB 14; Length 346;
Best Local Similarity 29.1%; Pred. No. 4.4e-25;
Matches 96; Conservative 63; Mismatches 164; Indels 7; Gaps 4;

Qy 21 IIAEIGNHOGDLDVAKRMIRMAKCGADCAKQKSELEFKNRKALE-RPYTSKHSWGK 79
Db 19 VVPEIGNHNGSLELAKIMVDAAFSTGAKIIKHQTHIVEDMSKAAKVIIPGNAKIS--- 75
Qy 80 TYGEKRLHLEFSDHQYRELQRYAEVGIFFTASGMDMAVEFLHNLNVPFFKVGSGDTNN 139
Db 76 -IYEIMQKCALDYKDELALKEYTEKGLVLTSTPFSRAGANRLEDMDGVSAPFKIGSGECNN 134
Qy 140 FPYLEKTAKGRPMVSSGMSQMDTMKVQIVKPLNPNFCFLOCTSAYPLOPEDVNLRV 199
Db 135 YPLIKHIAAFKPMIVSTGMNSIESIKPTVKILLDNEIPFVLMHTTNLYPTPHNLVRLNA 194
Qy 200 ISEYQKLFDPDIPGYSGHETGIAISVAALGAKVLERHITLDTKWGSDHSALEPGEL 259
Db 195 MLELKKEF-SCMVGLSDHTTDLNACLGAVAGACVLERHFTDSMHRSGPDIVCSMDTQAL 253
Qy 260 AEL-VRSVRLVERALGSPYKQLLPCEMACNEKLGKSVVAKVPIPEGTILTMWMLTVKVG 318
Db 254 KELLIOSEQMAIMRGNNESKAAKQEQVTIDFASFVSVSIKDIKKGEVLSMDNIWVKRPG 313
Qy 319 PKAYPPEDIENLVGKKVLTVEEDDTIMEE 348
Db 314 LGGISAAEFENILGKKALRDIENDTQLSYE 343

RESULT 12
US-10-303-128-35
; Sequence 35, Application US/10303128
; Publication No. US20030157656A1
; GENERAL INFORMATION:
; APPLICANT: Gilbert, Michel
; APPLICANT: Wakarchuk, Warren W.
; APPLICANT: National Research Council of Canada
; TITLE OF INVENTION: Campylobacter Glycosyltransferases for Biosynthesis of
; TITLE OF INVENTION: Gangliosides and Ganglioside Mimics
; FILE REFERENCE: 019633-000111US
; CURRENT APPLICATION NUMBER: US/10/303,128
; CURRENT FILING DATE: 2002-11-21

```

```

RESULT 11
US-10-303-118-35
; Sequence 35, Application US/10303118
; Publication No. US20030157655A1
; GENERAL INFORMATION:
; APPLICANT: Gilbert, Michel
; APPLICANT: Wakarchuk, Warren W.
; APPLICANT: National Research Council of Canada
; TITLE OF INVENTION: Campylobacter Glycosyltransferases for Biosynthesis of
; TITLE OF INVENTION: Gangliosides and Ganglioside Mimics
; FILE REFERENCE: 019633-000111US
; CURRENT APPLICATION NUMBER: US/10/303,118
; CURRENT FILING DATE: 2002-11-21
; PRIOR APPLICATION NUMBER: US/09/816,028
; PRIOR FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: US 60/118,213
; PRIOR FILING DATE: 1999-02-01
; PRIOR APPLICATION NUMBER: US 09/495,406
; PRIOR FILING DATE: 2000-01-31
; NUMBER OF SEQ ID NOS: 49
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 35
; LENGTH: 346
; TYPE: PRT
; ORGANISM: Campylobacter jejuni
; FEATURE:
; OTHER INFORMATION: sialic acid synthase from C. jejuni OH4384 (ORF 8a
; OTHER INFORMATION: of lipooligosaccharide (LOS) biosynthesis locus)
US-10-303-118-35

Query Match 17.8%; Score 336.5; DB 14; Length 346;
Best Local Similarity 29.1%; Pred. No. 4.4e-25;
Matches 96; Conservative 63; Mismatches 164; Indels 7; Gaps 4;

Qy 21 IIAEIGNHOGDLDVAKRMIRMAKCGADCAKQKSELEFKNRKALE-RPYTSKHSWGK 79
Db 19 VVPEIGNHNGSLELAKIMVDAAFSTGAKIIKHQTHIVEDMSKAAKVIIPGNAKIS--- 75
Qy 80 TYGEKRLHLEFSDHQYRELQRYAEVGIFFTASGMDMAVEFLHNLNVPFFKVGSGDTNN 139
Db 76 -IYEIMQKCALDYKDELALKEYTEKGLVLTSTPFSRAGANRLEDMDGVSAPFKIGSGECNN 134
Qy 140 FPYLEKTAKGRPMVSSGMSQMDTMKVQIVKPLNPNFCFLOCTSAYPLOPEDVNLRV 199
Db 135 YPLIKHIAAFKPMIVSTGMNSIESIKPTVKILLDNEIPFVLMHTTNLYPTPHNLVRLNA 194
Qy 200 ISEYQKLFDPDIPGYSGHETGIAISVAALGAKVLERHITLDTKWGSDHSALEPGEL 259
Db 195 MLELKKEF-SCMVGLSDHTTDLNACLGAVAGACVLERHFTDSMHRSGPDIVCSMDTQAL 253
Qy 260 AEL-VRSVRLVERALGSPYKQLLPCEMACNEKLGKSVVAKVPIPEGTILTMWMLTVKVG 318
Db 254 KELLIOSEQMAIMRGNNESKAAKQEQVTIDFASFVSVSIKDIKKGEVLSMDNIWVKRPG 313
Qy 319 PKAYPPEDIENLVGKKVLTVEEDDTIMEE 348
Db 314 LGGISAAEFENILGKKALRDIENDTQLSYE 343

RESULT 12
US-10-303-128-35
; Sequence 35, Application US/10303128
; Publication No. US20030157656A1
; GENERAL INFORMATION:
; APPLICANT: Gilbert, Michel
; APPLICANT: Wakarchuk, Warren W.
; APPLICANT: National Research Council of Canada
; TITLE OF INVENTION: Campylobacter Glycosyltransferases for Biosynthesis of
; TITLE OF INVENTION: Gangliosides and Ganglioside Mimics
; FILE REFERENCE: 019633-000111US
; CURRENT APPLICATION NUMBER: US/10/303,128
; CURRENT FILING DATE: 2002-11-21

```

```
; PRIOR APPLICATION NUMBER: US/09/816,028
; PRIOR FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: US 60/118,213
; PRIOR FILING DATE: 1999-02-01
; PRIOR APPLICATION NUMBER: US 09/495,406
; PRIOR FILING DATE: 2000-01-31
; NUMBER OF SEQ ID NOS: 49
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 35
; LENGTH: 346
; TYPE: PRP
; ORGANISM: Campylobacter jejuni
; FEATURE:
; OTHER INFORMATION: sialic acid synthase from C. jejuni OH4384 (ORF 8a)
; OTHER INFORMATION: of lipooligosaccharide (LOS) biosynthesis locus
US-10-303-128-35

Query Match      17.8%; Score 336.5; DB 14; Length 346;
Best Local Similarity 29.1%; Pred. No. 4.4e-25;
Matches 96; Conservative 63; Mismatches 164; Indels 7; Gaps 4;

QY 21 IIAEIGQNHQGDLDVAKRMIRMAKCEGADCAKFKSELEFKFNKALE-RPYTSKHSWGK 79
Db 19 VVPEIGINHGSLAKIMVDAAFSTGAKIIKHQTHIVEDMSKAAKKVIFGNAKIS--- 75
QY 80 TYGEHKHLEFSDQYRELQRYAEVGVIFFTASGDMEMAVEFLHNLNVPFFKVGSGDTNN 139
Db 76 -IYEMQKCALDYKDELALKEYTEKGLVYLSTPFSRAGANRLDMGVSFAFKIGSGECNN 134
QY 140 PYLEKTAAGRPVMISSGMSMDTMQVQIVKPLNPNFCFLOCTSAYPLOPEDVNLRV 199
Db 135 YPLIKHIAAFKKPMIVSTGMSISIKPTVKILLDNEIPFVLMTTNYPTPHNLRLNA 194
QY 200 ISEYQKLPDIPIGYSGHETGIAISVAALGAKVLERHITLDTKWGSDHSASLEPGEL 259
Db 195 MLEKKKEF-SCWVGLSDHTDNLACLGAVLACVLERHFTDSMHRSGPDIVCSMDTQAL 253
QY 260 AEL-VRSVRLVERALGSPKQLLPCCEMACNEKLGKSVVAKVPIEGTILTMDLTVKVG 318
Db 254 KELLIQEQWAIMRGNNESKAAKQEQVTIDFAFASVVSIDIKKGEVLSMDNIWVKRPG 313
QY 319 PKAYPPEDIFNLVGKVLVTVEEDDTIMEE 348
Db 314 LGGISAAEFENILGKALRIENDTQLSYE 343

RESULT 13
US-10-303-134-35
; Sequence 35, Application US/10303134
; Publication No. US20030157657A1
; GENERAL INFORMATION:
; APPLICANT: Gilbert, Michel
; APPLICANT: Wakarchuk, Warren W.
; APPLICANT: National Research Council of Canada
; TITLE OF INVENTION: Campylobacter Glycosyltransferases for Biosynthesis of
; TITLE OF INVENTION: Gangliosides and Ganglioside Mimics
; FILE REFERENCE: 019633-000111US
; CURRENT APPLICATION NUMBER: US/10/303,134
; CURRENT FILING DATE: 2002-11-21
; PRIOR APPLICATION NUMBER: US/09/816,028
; PRIOR FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: US 60/118,213
; PRIOR FILING DATE: 1999-02-01
; PRIOR APPLICATION NUMBER: US 09/495,406
; PRIOR FILING DATE: 2000-01-31
; NUMBER OF SEQ ID NOS: 49
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 35
; LENGTH: 346
; TYPE: PRP
; ORGANISM: Campylobacter jejuni
; FEATURE:
; OTHER INFORMATION: sialic acid synthase from C. jejuni OH4384 (ORF 8a)
; OTHER INFORMATION: of lipooligosaccharide (LOS) biosynthesis locus
US-10-303-134-35

Query Match      17.8%; Score 336.5; DB 14; Length 346;
Best Local Similarity 29.1%; Pred. No. 4.4e-25;
Matches 96; Conservative 63; Mismatches 164; Indels 7; Gaps 4;

QY 21 IIAEIGQNHQGDLDVAKRMIRMAKCEGADCAKFKSELEFKFNKALE-RPYTSKHSWGK 79
Db 19 VVPEIGINHGSLAKIMVDAAFSTGAKIIKHQTHIVEDMSKAAKKVIFGNAKIS--- 75
QY 80 TYGEHKHLEFSDQYRELQRYAEVGVIFFTASGDMEMAVEFLHNLNVPFFKVGSGDTNN 139
Db 76 -IYEMQKCALDYKDELALKEYTEKGLVYLSTPFSRAGANRLDMGVSFAFKIGSGECNN 134
QY 140 PYLEKTAAGRPVMISSGMSMDTMQVQIVKPLNPNFCFLOCTSAYPLOPEDVNLRV 199
Db 135 YPLIKHIAAFKKPMIVSTGMSISIKPTVKILLDNEIPFVLMTTNYPTPHNLRLNA 194
QY 200 ISEYQKLPDIPIGYSGHETGIAISVAALGAKVLERHITLDTKWGSDHSASLEPGEL 259
Db 195 MLEKKKEF-SCWVGLSDHTDNLACLGAVLACVLERHFTDSMHRSGPDIVCSMDTQAL 253
QY 260 AEL-VRSVRLVERALGSPKQLLPCCEMACNEKLGKSVVAKVPIEGTILTMDLTVKVG 318
Db 254 KELLIQEQWAIMRGNNESKAAKQEQVTIDFAFASVVSIDIKKGEVLSMDNIWVKRPG 313
QY 319 PKAYPPEDIFNLVGKVLVTVEEDDTIMEE 348
Db 314 LGGISAAEFENILGKALRIENDTQLSYE 343

RESULT 13
US-10-303-134-35
; Sequence 35, Application US/10303134
; Publication No. US20030157657A1
; GENERAL INFORMATION:
; APPLICANT: Gilbert, Michel
; APPLICANT: Wakarchuk, Warren W.
; APPLICANT: National Research Council of Canada
; TITLE OF INVENTION: Campylobacter Glycosyltransferases for Biosynthesis of
; TITLE OF INVENTION: Gangliosides and Ganglioside Mimics
; FILE REFERENCE: 019633-000111US
; CURRENT APPLICATION NUMBER: US/10/303,134
; CURRENT FILING DATE: 2002-11-21
; PRIOR APPLICATION NUMBER: US/09/816,028
; PRIOR FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: US 60/118,213
; PRIOR FILING DATE: 1999-02-01
; PRIOR APPLICATION NUMBER: US 09/495,406
; PRIOR FILING DATE: 2000-01-31
; NUMBER OF SEQ ID NOS: 49
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 35
; LENGTH: 346
; TYPE: PRP
; ORGANISM: Campylobacter jejuni
; FEATURE:
; OTHER INFORMATION: sialic acid synthase from C. jejuni OH4384 (ORF 8a)
```

```
; OTHER INFORMATION: of lipooligosaccharide (LOS) biosynthesis locus)
US-10-303-134-35

Query Match      17.8%; Score 336.5; DB 14; Length 346;
Best Local Similarity 29.1%; Pred. No. 4.4e-25;
Matches 96; Conservative 63; Mismatches 164; Indels 7; Gaps 4;

QY 21 IIAEIGQNHQGDLDVAKRMIRMAKCEGADCAKFKSELEFKFNKALE-RPYTSKHSWGK 79
Db 19 VVPEIGINHGSLAKIMVDAAFSTGAKIIKHQTHIVEDMSKAAKKVIFGNAKIS--- 75
QY 80 TYGEHKHLEFSDQYRELQRYAEVGVIFFTASGDMEMAVEFLHNLNVPFFKVGSGDTNN 139
Db 76 -IYEMQKCALDYKDELALKEYTEKGLVYLSTPFSRAGANRLDMGVSFAFKIGSGECNN 134
QY 140 PYLEKTAAGRPVMISSGMSMDTMQVQIVKPLNPNFCFLOCTSAYPLOPEDVNLRV 199
Db 135 YPLIKHIAAFKKPMIVSTGMSISIKPTVKILLDNEIPFVLMTTNYPTPHNLRLNA 194
QY 200 ISEYQKLPDIPIGYSGHETGIAISVAALGAKVLERHITLDTKWGSDHSASLEPGEL 259
Db 195 MLEKKKEF-SCWVGLSDHTDNLACLGAVLACVLERHFTDSMHRSGPDIVCSMDTQAL 253
QY 260 AEL-VRSVRLVERALGSPKQLLPCCEMACNEKLGKSVVAKVPIEGTILTMDLTVKVG 318
Db 254 KELLIQEQWAIMRGNNESKAAKQEQVTIDFAFASVVSIDIKKGEVLSMDNIWVKRPG 313
QY 319 PKAYPPEDIFNLVGKVLVTVEEDDTIMEE 348
Db 314 LGGISAAEFENILGKALRIENDTQLSYE 343

RESULT 14
US-10-303-162-35
; Sequence 35, Application US/10303162
; Publication No. US20030157658A1
; GENERAL INFORMATION:
; APPLICANT: Gilbert, Michel
; APPLICANT: Wakarchuk, Warren W.
; APPLICANT: National Research Council of Canada
; TITLE OF INVENTION: Campylobacter Glycosyltransferases for Biosynthesis of
; TITLE OF INVENTION: Gangliosides and Ganglioside Mimics
; FILE REFERENCE: 019633-000111US
; CURRENT APPLICATION NUMBER: US/10/303,162
; CURRENT FILING DATE: 2002-11-21
; PRIOR APPLICATION NUMBER: US/09/816,028
; PRIOR FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: US 60/118,213
; PRIOR FILING DATE: 1999-02-01
; PRIOR APPLICATION NUMBER: US 09/495,406
; PRIOR FILING DATE: 2000-01-31
; NUMBER OF SEQ ID NOS: 49
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 35
; LENGTH: 346
; TYPE: PRP
; ORGANISM: Campylobacter jejuni
; FEATURE:
; OTHER INFORMATION: sialic acid synthase from C. jejuni OH4384 (ORF 8a)
; OTHER INFORMATION: of lipooligosaccharide (LOS) biosynthesis locus
US-10-303-162-35

Query Match      17.8%; Score 336.5; DB 14; Length 346;
Best Local Similarity 29.1%; Pred. No. 4.4e-25;
Matches 96; Conservative 63; Mismatches 164; Indels 7; Gaps 4;

QY 21 IIAEIGQNHQGDLDVAKRMIRMAKCEGADCAKFKSELEFKFNKALE-RPYTSKHSWGK 79
Db 19 VVPEIGINHGSLAKIMVDAAFSTGAKIIKHQTHIVEDMSKAAKKVIFGNAKIS--- 75
QY 80 TYGEHKHLEFSDQYRELQRYAEVGVIFFTASGDMEMAVEFLHNLNVPFFKVGSGDTNN 139
Db 76 -IYEMQKCALDYKDELALKEYTEKGLVYLSTPFSRAGANRLDMGVSFAFKIGSGECNN 134
QY 140 PYLEKTAAGRPVMISSGMSMDTMQVQIVKPLNPNFCFLOCTSAYPLOPEDVNLRV 199
Db 135 YPLIKHIAAFKKPMIVSTGMSISIKPTVKILLDNEIPFVLMTTNYPTPHNLRLNA 194
QY 200 ISEYQKLPDIPIGYSGHETGIAISVAALGAKVLERHITLDTKWGSDHSASLEPGEL 259
Db 195 MLEKKKEF-SCWVGLSDHTDNLACLGAVLACVLERHFTDSMHRSGPDIVCSMDTQAL 253
QY 260 AEL-VRSVRLVERALGSPKQLLPCCEMACNEKLGKSVVAKVPIEGTILTMDLTVKVG 318
Db 254 KELLIQEQWAIMRGNNESKAAKQEQVTIDFAFASVVSIDIKKGEVLSMDNIWVKRPG 313
QY 319 PKAYPPEDIFNLVGKVLVTVEEDDTIMEE 348
Db 314 LGGISAAEFENILGKALRIENDTQLSYE 343

RESULT 14
US-10-303-162-35
; Sequence 35, Application US/10303162
; Publication No. US20030157658A1
; GENERAL INFORMATION:
; APPLICANT: Gilbert, Michel
; APPLICANT: Wakarchuk, Warren W.
; APPLICANT: National Research Council of Canada
; TITLE OF INVENTION: Campylobacter Glycosyltransferases for Biosynthesis of
; TITLE OF INVENTION: Gangliosides and Ganglioside Mimics
; FILE REFERENCE: 019633-000111US
; CURRENT APPLICATION NUMBER: US/10/303,162
; CURRENT FILING DATE: 2002-11-21
; PRIOR APPLICATION NUMBER: US/09/816,028
; PRIOR FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: US 60/118,213
; PRIOR FILING DATE: 1999-02-01
; PRIOR APPLICATION NUMBER: US 09/495,406
; PRIOR FILING DATE: 2000-01-31
; NUMBER OF SEQ ID NOS: 49
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 35
; LENGTH: 346
; TYPE: PRP
; ORGANISM: Campylobacter jejuni
; FEATURE:
; OTHER INFORMATION: sialic acid synthase from C. jejuni OH4384 (ORF 8a)
; OTHER INFORMATION: of lipooligosaccharide (LOS) biosynthesis locus
US-10-303-162-35

Query Match      17.8%; Score 336.5; DB 14; Length 346;
Best Local Similarity 29.1%; Pred. No. 4.4e-25;
Matches 96; Conservative 63; Mismatches 164; Indels 7; Gaps 4;

QY 21 IIAEIGQNHQGDLDVAKRMIRMAKCEGADCAKFKSELEFKFNKALE-RPYTSKHSWGK 79
Db 19 VVPEIGINHGSLAKIMVDAAFSTGAKIIKHQTHIVEDMSKAAKKVIFGNAKIS--- 75
QY 80 TYGEHKHLEFSDQYRELQRYAEVGVIFFTASGDMEMAVEFLHNLNVPFFKVGSGDTNN 139
Db 76 -IYEMQKCALDYKDELALKEYTEKGLVYLSTPFSRAGANRLDMGVSFAFKIGSGECNN 134
QY 140 PYLEKTAAGRPVMISSGMSMDTMQVQIVKPLNPNFCFLOCTSAYPLOPEDVNLRV 199
Db 135 YPLIKHIAAFKKPMIVSTGMSISIKPTVKILLDNEIPFVLMTTNYPTPHNLRLNA 194
QY 200 ISEYQKLPDIPIGYSGHETGIAISVAALGAKVLERHITLDTKWGSDHSASLEPGEL 259
Db 195 MLEKKKEF-SCWVGLSDHTDNLACLGAVLACVLERHFTDSMHRSGPDIVCSMDTQAL 253
QY 260 AEL-VRSVRLVERALGSPKQLLPCCEMACNEKLGKSVVAKVPIEGTILTMDLTVKVG 318
Db 254 KELLIQEQWAIMRGNNESKAAKQEQVTIDFAFASVVSIDIKKGEVLSMDNIWVKRPG 313
QY 319 PKAYPPEDIFNLVGKVLVTVEEDDTIMEE 348
Db 314 LGGISAAEFENILGKALRIENDTQLSYE 343
```

Search completed: September 13, 2004, 14:55:33
Job time : 132 secs

```
Qy 140 PPYLEKTAAGRPVWISSGQSMQMTMKQVQIVKPLNPNFCFLQCTSAYPELPQEDVNLRV 199
Db 135 YPLIKHIAAFKKMIVSTGMNIESIKPTVKILLDNEIPFVLMHTTNLYPTPHNLVRLNA 194
Qy 200 ISEYQKLFDPDIPICYSGHETGIAISVAALGAKVLERHITLDTKWTGSDHSASLEPGEL 259
Db 195 MLELKKEF-SQWVGLSDHTDNLACLGAVALGACVLERHFTDSMRSGPDIIVCSMDTQAL 253
Qy 260 AEL-VRSVRLVERALGSPTKQLLPCMACNEKLGKSVVAKVPIEGTILTMMDMLTVKVGE 318
Db 254 KELIQSEQMAIMGNESKAAKQEQVTIDFASFASVVISIKDKKGEVLSDNIWVKRPG 313
Qy 319 PKAYPPEDIFNLGVKKVILVVEEDDTIMEE 348
Db 314 LGGISAAEFENILGKKALRDIENDTOLSYE 343
```

RESULT 15

```
US-09-882-227-490
; Sequence 490, Application US/09882227
; Publication No. US20030158396A1
; GENERAL INFORMATION:
; APPLICANT: Kleanthous, Harold
; APPLICANT: Al-Garawi, Amal
; APPLICANT: Miller, Charles
; APPLICANT: Tomb, Jean-Francois
; APPLICANT: Ooomen, Raymond P.
; TITLE OF INVENTION: Identification of Polynucleotides
; TITLE OF INVENTION: Encoding No. US20030158396A1el Helicobacter Polypeptides in the
; FILE REFERENCE: 06132/047002
; CURRENT APPLICATION NUMBER: US/09/882,227
; CURRENT FILING DATE: 2001-06-15
; PRIOR APPLICATION NUMBER: US 08/902,615
; PRIOR FILING DATE: 1997-07-29
; NUMBER OF SEQ ID NOS: 638
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 490
; LENGTH: 340
; TYPE: PRT
; ORGANISM: Helicobacter pylori
US-09-882-227-490
```

Query Match 16.1%; Score 303; DB 10; Length 340;
Best Local Similarity 27.7%; Pred. No. 1.1e-21;
Matches 91; Conservative 66; Mismatches 149; Indels 22; Gaps 10;

```
Qy 18 PCFIIAEIGQNHQGDLDVAKMIRMAKECGADCAKFOK-----SELEKFNKALEREPT 72
Db 4 PPKIVAEISANHNQDLNAKESLHAIKSGADFVKLQTYTPTSCMTLNSKEDPFIQGTLM 63
Qy 73 SKHSGKTYGSHKHLEFSDHQYRELQRYABEVGIFTASGMDMAVEFLHNLNVPFEKV 132
Db 64 DKNELYLYQXASTPLEW-HAELFELARKL-DLGIF--SSPSSQALELLESNCMPYKI 119
Qy 133 GSGDTNPNFYLEKTAAGRPVWISSGQSMQMTMKQVQIVKPLNPNF--CFLOCTSAAYPL 190
Db 120 ASFEIVDLDLIEKAARTQKPIILSSGIATHTLQDAISLCRRVN-NFDITLLKCVSAYPS 178
Qy 191 QPEDVNLRISEYQKLFDPDIPIGYGHETGTATISVAALGAKVLERHITLDTKWTGSDH 250
Db 179 KIEDANLLSMVKLGHIF-GVKFGLSDHTIGSLCPILATTLGASMTIERFILNKSLOTPTS 237
Qy 251 SASLEPGELAEVRLVERALGSPTKQLLPCMACNEKLGKSVVAKVPIEGTILT-- 308
Db 238 AFSMDPENGFKSNVEAIKQSVLALGEEPRINPTLEKRRFFARSILFVKIDIOKGEALTEN 297
Qy 309 -MDMLTVKVG-EPKAYPPEDIFNLVGKK 334
Db 298 NIKALRPNLGLHLPKFKYK-----EILGQK 320
```

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 13, 2004, 14:23:23 ; Search time 126 Seconds
(without alignments)
805.036 Million cell updates/sec

Title: US-10-759-277-4
Perfect score: 1887
Sequence: 1 MPLELELCPRWVGQHPGCP.....EEDDTIMEELVDNHGKIKS 359

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_29Jan04.*
1: Geneseq1980s.*
2: Geneseq1990s.*
3: Geneseq2000s.*
4: Geneseq2001s.*
5: Geneseq2002s.*
6: Geneseq2003as.*
7: Geneseq2003bs.*
8: Geneseq2004s.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	1887	100.0	359	3	AY96101	Human sia
2	1887	100.0	359	3	AY90352	Human gly
3	1887	100.0	359	4	AB84683	Amino aci
4	1887	100.0	359	6	AA026546	Human SA
5	1883	99.8	359	5	AB90221	Human pol
6	1880	99.6	359	4	AA03986	Human pol
7	1876	99.4	359	4	AB93183	Human pro
8	1789	94.8	400	4	AA41172	Human pol
9	783	41.5	338	4	AB71364	Drosophila
10	544.5	28.9	338	3	AY68965	Cps2p pro
11	532.5	28.2	341	5	ABP26810	Streptoco
12	510.5	27.1	346	3	AY96102	Escherich
13	510.5	27.1	346	4	AB84684	Amino aci
14	510.5	27.1	346	6	AA026547	Bacterial
15	419.5	22.2	125	4	ABG17251	Novel hum
16	395	20.9	153	4	ABG27650	Novel hum
17	382.5	20.3	349	5	AA49716	N. mening
18	381	20.2	123	4	AA076152	Human col
19	336.5	17.8	346	6	AB197207	Campyloba
20	336.5	17.8	346	6	AB18497	Campyloba
21	320.5	17.0	328	5	AB78389	Amino aci
22	303	16.1	340	2	AAW98812	H. pylori
23	266.5	14.1	136	4	ABG27649	Novel hum
24	159.5	8.5	201	5	ABU51188	Helicobac
25	149.5	7.9	182	5	ABU50662	Helicobac

26	113.5	6.0	333	6	ABU29572	Protein e
27	113.5	6.0	343	7	ADC97084	E. faeciu
28	110	5.8	265	4	AAB96139	Putative
29	108.5	5.7	633	6	ABU35582	Protein e
30	107.5	5.7	337	6	ABU23592	Protein e
31	107	5.7	330	6	ABU39673	Protein e
32	104.5	5.5	769	6	ABM70636	Phototrab
33	102.5	5.4	610	6	ADB06924	Alloiococ
34	102.5	5.4	629	6	ADB06926	Alloiococ
35	102.5	5.4	630	6	ADB06928	Alloiococ
36	101.5	5.4	337	6	ABU25380	Protein e
37	101.5	5.4	341	6	ABU29304	Protein e
38	101	5.4	66	7	ADC79295	Z. elonga
39	101	5.4	350	7	ADC59279	Comamonas
40	101	5.4	350	7	ADC52503	Testoster
41	100	5.3	65	7	ADC79293	Z. elonga
42	100	5.3	547	6	ADA12178	Acinetoba
43	100	5.3	547	7	ADC61133	Baeyer-Vi
44	100	5.3	547	7	AAE39407	Acinetoba
45	100	5.3	1109	6	ABU71017	Human adi

ALIGNMENTS

RESULT 1
AY96101
ID AY96101 standard; protein; 359 AA.

XX AC AY96101;

XX DT 19-DEC-2000 (first entry)

XX DE Human sialic acid synthetase.

XX KW Sialic acid synthetase; human; sas gene; sialylation; glycoprotein;
plasminogen; transferrin; thyrotropin; Na+,K+-ATPase.

XX OS Homo sapiens.

XX PN WO200052135-A2.

XX PD 08-SEP-2000.

XX PF 01-MAR-2000; 2000WO-US005313.

XX PR 02-MAR-1999; 99US-0122582P.

XX PR 08-DEC-1999; 99US-0169624P.

XX (HUMA-) HUMAN GENOME SCI INC.
(UYJO) UNIV JOHNS HOPKINS.

XX (UYWY-) UNIV WYOMING.

XX Betenbaugh MJ, Lawrence S, Lee YC, Jarvis D, Coleman TA;

XX WPI; 2000-572178/53.

XX N-P8DB; AA50569.

XX Recombinant production of sialylated glycoproteins using cells in which
the expression of enzymes, e.g. sialic acid synthetase, involved in the
sialylation reaction has been altered.

XX Claim 16; Page 105-106; 144pp; English.

XX The present sequence is that of human sialic acid synthetase (SAS), an
enzyme that condenses Mann6P or Man-6-P with UDP to form Neu5Ac and
KDN phosphates, respectively. The sequence was deduced from SAS cDNA (see
AA50569). Northern blots indicated ubiquitous transcription of the SAS
gene in a selection of tissues. The invention provides methods and
recombinantly engineered cells for producing glycoproteins having
sialylated oligosaccharides. The methods involve altering the expression
of enzymes involved in carbohydrate processing. A claimed cell producing
sialylated glycoprotein at above endogenous levels expresses at least 1

CC (preferably human) enzyme selected from GlcNAc-2 epimerase, an enzyme
 CC catalyzing the conversion of UDP-GlcNAc to MannNAc, sialic acid
 CC synthetase, aldolase, CMP-sialic acid synthetase and CMP-sialic acid
 CC transporter at above endogenous levels. Endogenous N-
 CC acetylglucosaminidase activity may be suppressed. A claimed method for
 CC manipulating glycoprotein in an insect cell comprises enhancing the
 CC expression of 1 of the above enzymes, and a claimed method for producing
 CC sialylated glycoproteins involves expressing a heterologous protein
 CC (especially plasminogen, transferrin, Na⁺K⁺-ATPase or thyrotropin) in
 CC the insect cell. Yeast, insect, fungal, plant and bacterial host cells
 CC can be engineered to produce new forms of sialylated glycoproteins,
 CC higher concentrations of sialylated glycoproteins and/or elevated
 CC concentrations of donor substrates (e.g. nucleotide sugars) required for
 CC sialylation
 XX
 SQ Sequence 359 AA;

Query Match 100.0%; Score 1887; DB 3; Length 359;
 Best Local Similarity 100.0%; Pred. No. 9.4e-184;
 Matches 359; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MPELELCGRWVGQHPFCFIIAETGQNHQGLDVAKRMIRMAKCGADCAKFKQSELEF 60
 Db 1 MPELELCGRWVGQHPFCFIIAETGQNHQGLDVAKRMIRMAKCGADCAKFKQSELEF 60
 QY 61 KFNKALERPPTS KSHSGKTYGEHKHLEFSDQYRELQRYAEVEVGIFFTAGMDMAVE 120
 Db 61 KFNKALERPPTS KSHSGKTYGEHKHLEFSDQYRELQRYAEVEVGIFFTAGMDMAVE 120
 QY 121 FLHNLNVPFFKVGSGDTNNFPYLEKTAKKGRPMWISSGQSMQMDTKQYQIVKPLNPFC 180
 Db 121 FLHNLNVPFFKVGSGDTNNFPYLEKTAKKGRPMWISSGQSMQMDTKQYQIVKPLNPFC 180
 QY 181 FLQCTSAYPELOPEDVNLRVISEYOKLPDIPIGYSGHETGIAISVAALGAKVLERHIT 240
 Db 181 FLQCTSAYPELOPEDVNLRVISEYOKLPDIPIGYSGHETGIAISVAALGAKVLERHIT 240
 QY 241 LDKTWKGSDHSASLEPGELAEVLRSVRLVERALGSPTKQLLPCEMACNEKLGKSVVAVK 300
 Db 241 LDKTWKGSDHSASLEPGELAEVLRSVRLVERALGSPTKQLLPCEMACNEKLGKSVVAVK 300
 QY 301 IPEGTILTMDLTVKVGEPKAYPPEDI FNLVGKKVLTVVEEDDTIMEELVDNHGKKIKS 359
 Db 301 IPEGTILTMDLTVKVGEPKAYPPEDI FNLVGKKVLTVVEEDDTIMEELVDNHGKKIKS 359

RESULT 2
 AAY90352
 ID AAY90352 standard; protein; 359 AA.

XX AAY90352;

XX 04-DEC-2000 (first entry)

DE Human glycosylation enzyme clone HASAA37 protein sequence.

XX Human; glycosylation enzyme; glycolysis; myoglobinuria; tumour marker;
 KW immunotherapy; cosmetic surgery; metabolism; immune system disorder;
 KW haematopoietic cell deficiency; blood coagulation disorder; asthma;
 KW afibrinogenemia; blood platelet disorder; thrombocytopenia; purpura;
 KW autoimmune disorder; Addison's disease; multiple sclerosis; purpura;
 KW allergic encephalomyelitis; allergic reaction; organ rejection;
 KW graft-versus-host disease; inflammation; hyperproliferative disorder;
 KW sarcoidosis; infection; gene therapy; CMP sialic acid synthetase.

XX Homo sapiens.

OS WO200052136-A2.

XX 08-SEP-2000.

XX 01-MAR-2000; 2000WO-US005325.

PR 02-MAR-1999; 99US-0122409P.
 XX (HUMA-) HUMAN GENOME SCI INC.
 PA Coleman TA;
 XX WPI; 2000-572179/53.
 DR N-PSDB; AAA37763.

XX New human glycosylation enzymes cytidine 5'-monophosphate sialic acid
 PT synthetase, sialic acid synthetase and aldolase and nucleic acids
 PT encoding the proteins for treating e.g., immune system disorders,
 PT microbial diseases.

XX Claim 12; Page 110-111; 115pp; English.

CC This sequence represents a human glycosylation enzyme clone of the
 CC invention, designated HASAA37. This protein clone is a sialic acid
 CC synthetase. The sequences are useful as reagents for the differential
 CC identification of the tissues or cell types present in a biological
 CC sample, as immunological probes, for treating a disease or condition
 CC resulting from under expression of such polypeptide, for the detection
 CC and/or treatment of disorders involving aberrant glycolysis, e.g. cramps,
 CC myoglobinuria, and as tumour marker and/or immunotherapy targets. They
 CC may also be used to differentiate, proliferate and attract cells leading
 CC to the regeneration of tissues, to modulate mammalian characteristics
 CC (e.g. in cosmetic surgery) or mammalian metabolism affecting catabolism,
 CC anabolism processing, utilisation and energy storage, to change a
 CC mammal's mental state by influencing and as a food additive or
 CC preservative. The proteins can be used to assay protein levels in a
 CC sample, as a marker or detector of an immune system disorder, to inhibit
 CC cytokine activity, and as a vaccine. They may further be used to treat
 CC immune system or of haematopoietic cell deficiencies or disorders, blood
 CC coagulation disorders (e.g. afibrinogenemia), blood platelet disorders
 CC (e.g. thrombocytopenia), wounds resulting from trauma or surgery,
 CC autoimmune disorders (e.g. Addison's disease, multiple sclerosis,
 CC allergic encephalomyelitis), allergic reactions (e.g. asthma), organ
 CC rejection, graft-versus-host disease, inflammation, hyperproliferative
 CC disorders (e.g. neoplasia, purpura, sarcoidosis), diseases caused by
 CC viruses (e.g. hepatitis, meningitis, AIDS), bacteria and fungi (include
 CC e.g. tuberculosis, conjunctivitis, sepsis, typhoid, chlamydia,
 CC cellulitis), and diseases caused by parasites (e.g. amoebiasis,
 CC coccidiosis, leishmaniasis, scabies, malaria, toxoplasmosis)

XX Sequence 359 AA;

Query Match 100.0%; Score 1887; DB 3; Length 359;
 Best Local Similarity 100.0%; Pred. No. 9.4e-184;
 Matches 359; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MPELELCGRWVGQHPFCFIIAETGQNHQGLDVAKRMIRMAKCGADCAKFKQSELEF 60
 Db 1 MPELELCGRWVGQHPFCFIIAETGQNHQGLDVAKRMIRMAKCGADCAKFKQSELEF 60
 QY 61 KFNKALERPPTS KSHSGKTYGEHKHLEFSDQYRELQRYAEVEVGIFFTAGMDMAVE 120
 Db 61 KFNKALERPPTS KSHSGKTYGEHKHLEFSDQYRELQRYAEVEVGIFFTAGMDMAVE 120
 QY 121 FLHNLNVPFFKVGSGDTNNFPYLEKTAKKGRPMWISSGQSMQMDTKQYQIVKPLNPFC 180
 Db 121 FLHNLNVPFFKVGSGDTNNFPYLEKTAKKGRPMWISSGQSMQMDTKQYQIVKPLNPFC 180
 QY 181 FLQCTSAYPELOPEDVNLRVISEYOKLPDIPIGYSGHETGIAISVAALGAKVLERHIT 240
 Db 181 FLQCTSAYPELOPEDVNLRVISEYOKLPDIPIGYSGHETGIAISVAALGAKVLERHIT 240
 QY 241 LDKTWKGSDHSASLEPGELAEVLRSVRLVERALGSPTKQLLPCEMACNEKLGKSVVAVK 300
 Db 241 LDKTWKGSDHSASLEPGELAEVLRSVRLVERALGSPTKQLLPCEMACNEKLGKSVVAVK 300
 QY 301 IPEGTILTMDLTVKVGEPKAYPPEDI FNLVGKKVLTVVEEDDTIMEELVDNHGKKIKS 359
 Db 301 IPEGTILTMDLTVKVGEPKAYPPEDI FNLVGKKVLTVVEEDDTIMEELVDNHGKKIKS 359

RESULT 3
AAB84683
ID AAB84683 standard; protein; 359 AA.
XX AAB84683;
AC AAB84683;
XX 17-SEP-2001 (first entry)
DT 17-SEP-2001 (first entry)
XX Amino acid sequence of a human sialic acid synthetase.
DE Carbohydrate processing; sialylation; sialylated glycoprotein; CMP-SA;
XX cytidine monophosphate-sialic acid; N-acetylglucosamine-2 epimerase;
KW sialic acid synthetase; aldolase; CMP-SA synthetase; CMP-SA transporter;
KW vaccine.
XX Homo sapiens.
OS Homo sapiens.
XX WO200142492-A1.
FN WO200142492-A1.
XX 14-JUN-2001.
XX 07-DEC-2000; 2000WO-US033136.
XX 09-DEC-1999; 99US-0169899P.
XX (HUMA-) HUMAN GENOME SCI INC.
PA (UYJO) UNIV JOHNS HOPKINS.
PA (UTEM) UNIV TEMPLE.
PA (UYWY-) UNIV WYOMING.
XX Betenbaugh MJ, Lawrence S, Lee YC, Coleman TA, Palter K;
PI Jarvis D;
XX WPI; 2001-441575/47.
DR N-PSDB; AAB84683.
XX Cells producing cytidine monophosphate-sialic acid and sialylated
PT glycoprotein above endogenous levels for production of vaccines and
PT therapeutics.
PS Claim 16; Fig 32; 182pp; English.
XX The specification describes a method for manipulating carbohydrate
CC processing pathways in cells of interest. The methods are used to
CC manipulate multiple pathways involved with the sialylation reaction by
CC using recombinant DNA technology and substrate feeding approaches to
CC enable the production of sialylated glycoproteins in the cells. The
CC sialylation process involves the post-translational addition of the donor
CC substrate cytidine monophosphate-sialic acid (CMP-SA) onto a specific
CC acceptor carbohydrate. The cells express at least one enzyme, selected
CC from N-acetylglucosamine-2 epimerase, sialic acid synthetase, aldolase,
CC CMP-SA synthetase and CMP-SA transporter, above endogenous levels. The
CC cells are useful for producing complex sialylated glycoproteins in cells
CC of interest, especially insect cells. Glycoproteins containing sialylated
CC oligosaccharides are useful as vaccines, therapeutics and diagnostic
CC tools. Cells producing complex sialylated glycoproteins are useful for
CC enhancing the value of heterologous expression systems and increasing the
CC application of heterologous cell expression products as vaccines,
CC therapeutics and diagnostic tools as well as increasing the variety of
CC heterologous proteins that can be produced and lowering biotechnology
CC production costs. The present sequence represents a human sialic acid
CC synthetase, which is used in the method of the invention
XX Sequence 359 AA;
SQ Query Match 100.0%; Score 1887; DB 4; Length 359;
Best Local Similarity 100.0%; Pred. No. 9,4e-184;
Matches 359; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MPLELELCGRWVGQHPCEIIAEIQNHQDLDVAKEMIRVAKECGADCAKFOKSELEF 60
|||

CC manipulating glycoprotein production in an insect cell. Further methods
CC of the invention are useful for producing sialylated glycoprotein. The
CC sialylated glycoprotein produced by the above mentioned methods are
CC useful as pharmaceutical compositions, vaccines, diagnostics and
CC therapeutics. This sequence represents the human SA synthetase protein of
CC the invention
XX
SQ Sequence 359 AA;

Query Match 100.0%; Score 1887; DB 6; Length 359;
Best Local Similarity 100.0%; Pred. No. 9.4e-184;
Matches 359; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX

QY 1 MPELELCPRWVGQHPFCFIIAIGQNHQGDLDVAKRMIRWAKGADCAKFKSELEF 60
DB 1 MPELELCPRWVGQHPFCFIIAIGQNHQGDLDVAKRMIRWAKGADCAKFKSELEF 60
QY 61 KFNKALERPPTYSKHSWGKTYGHEKRLHLEFSDQYRELQRYAEVGIFFTAGMDMAVE 120
DB 61 KFNKALERPPTYSKHSWGKTYGHEKRLHLEFSDQYRELQRYAEVGIFFTAGMDMAVE 120
QY 121 FLHELNVFPFKVSGDNTNFPVLEKTAQKGRPMWISSGQSDMTMKQVQIVKPLNPNFC 180
DB 121 FLHELNVFPFKVSGDNTNFPVLEKTAQKGRPMWISSGQSDMTMKQVQIVKPLNPNFC 180
QY 181 FLQCTSAVLPQEDVNLRVISEYQKLFDPDIPIGYSGHETGIAISVAALGAKVLERHIT 240
DB 181 FLQCTSAVLPQEDVNLRVISEYQKLFDPDIPIGYSGHETGIAISVAALGAKVLERHIT 240
QY 241 LDKTWKSDHSASLEPGLAEIVRSVRLVERALGSPKQLLPCMACNEKLGKSVVAVK 300
DB 241 LDKTWKSDHSASLEPGLAEIVRSVRLVERALGSPKQLLPCMACNEKLGKSVVAVK 300
QY 301 IPEGTILTMDLTVKVGEPKAYPPEDI FNLVGKKVLVTVEEDDTIMEELVDNHGKKIKS 359
DB 301 IPEGTILTMDLTVKVGEPKAYPPEDI FNLVGKKVLVTVEEDDTIMEELVDNHGKKIKS 359

RESULT 5
ABB90221 ID ABB90221 standard; protein; 359 AA.
XX AC ABB90221;
XX DT 24-MAY-2002 (first entry)
XX DE Human polypeptide SEQ ID NO 2597.
XX KW Cytostatic; immunosuppressive; nontropic; neuroprotective; antiviral;
KW antiallergic; hepatotropic; antidiabetic; antiinflammatory; antiulcer;
KW vulnary; anticonvulsant; antibacterial; antifungal; antiparasitic;
KW cardiant; gene therapy; cancer; immune disorder; cardiovascular disorder;
KW neurological disease; infection; human; secreted protein.
XX OS Homo sapiens.
XX PN WO200190304-A2.
XX PD 29-NOV-2001.
XX PF 18-MAY-2001; 2001WO-US016450.
XX PR 19-MAY-2000; 2000US-0205515P.
XX PA (HUMA-) HUMAN GENOME SCI INC.
XX PI Birse CE, Rosen CA;
XX DR WPI: 2002-122018/16.
XX DR N-PSDB; ABL90630.
XX PT Novel 1405 isolated polypeptides, useful for diagnosis, treatment and
PT prevention of neural, immune system, muscular, reproductive,

PT gastrointestinal, pulmonary, cardiovascular, renal and proliferative
PT disorders.
XX

PS Claim 11; SEQ ID NO 2597; 2081pp + Sequence Listing; English.

XX The invention relates to novel genes (ABL89449-ABL90853) and proteins
CC (ABB9040-ABB9044) useful for preventing, treating or ameliorating
CC medical conditions e.g. by protein or gene therapy. The genes are
CC isolated from a range of human tissues disclosed in the specification.
CC The nucleic acids, proteins, antibodies and (ant)agonists are useful in
CC the diagnosis, treatment and prevention of: (a) cancer, e.g. breast and
CC ovarian cancer and other cancers of the adrenal gland, bone, bone marrow,
CC breast, gastrointestinal tract, liver, lung, or urogenital; (b) immune
CC disorders e.g. Addison's disease, allergies, autoimmune haemolytic
CC anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's disease,
CC multiple sclerosis, rheumatoid arthritis and ulcerative colitis; (c)
CC cardiovascular disorders such as myocardial ischaemia; (d) wound healing
CC; (e) neurological diseases e.g. cerebral anoxia and epilepsy; and (f)
CC infectious diseases such as viral, bacterial, fungal and parasitic
CC infections. Note: The sequence data for this patent did not form part of
CC the printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 359 AA;

Query Match 99.8%; Score 1883; DB 5; Length 359;
Best Local Similarity 99.7%; Pred. No. 2.4e-183;
Matches 358; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MPELELCPRWVGQHPFCFIIAIGQNHQGDLDVAKRMIRWAKGADCAKFKSELEF 60
DB 1 MPELELCPRWVGQHPFCFIIAIGQNHQGDLDVAKRMIRWAKGADCAKFKSELEF 60
QY 61 KFNKALERPPTYSKHSWGKTYGHEKRLHLEFSDQYRELQRYAEVGIFFTAGMDMAVE 120
DB 61 KFNKALERPPTYSKHSWGKTYGHEKRLHLEFSDQYRELQRYAEVGIFFTAGMDMAVE 120
QY 121 FLHELNVFPFKVSGDNTNFPVLEKTAQKGRPMWISSGQSDMTMKQVQIVKPLNPNFC 180
DB 121 FLHELNVFPFKVSGDNTNFPVLEKTAQKGRPMWISSGQSDMTMKQVQIVKPLNPNFC 180
QY 181 FLQCTSAVLPQEDVNLRVISEYQKLFDPDIPIGYSGHETGIAISVAALGAKVLERHIT 240
DB 181 FLQCTSAVLPQEDVNLRVISEYQKLFDPDIPIGYSGHETGIAISVAALGAKVLERHIT 240
QY 241 LDKTWKSDHSASLEPGLAEIVRSVRLVERALGSPKQLLPCMACNEKLGKSVVAVK 300
DB 241 LDKTWKSDHSASLEPGLAEIVRSVRLVERALGSPKQLLPCMACNEKLGKSVVAVK 300
QY 301 IPEGTILTMDLTVKVGEPKAYPPEDI FNLVGKKVLVTVEEDDTIMEELVDNHGKKIKS 359
DB 301 IPEGTILTMDLTVKVGEPKAYPPEDI FNLVGKKVLVTVEEDDTIMEELVDNHGKKIKS 359

RESULT 6
AAM39986 ID AAM39986 standard; protein; 359 AA.
XX AC AAM39986;
XX DT 22-OCT-2001 (first entry)
XX DE Human polypeptide SEQ ID NO 3131.
XX

XX Human; nontropic; immunosuppressant; cytostatic; gene therapy; cancer;
KW peripheral nervous system; neuropathy; central nervous system; CNS;
KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
KW leukaemia.
XX OS Homo sapiens.
XX

QY 1 MPELELCPRWVGQHPFCFIIAETGQNHQGDLDVAKRMIRMAKEGADCAKFKQSELEF 60
Db 1 MPELELCPRWVGQHPFCFIIAETGQNHQGDLDVAKRMIRMAKEGADCAKFKQSELEF 60
QY 61 KFNKALERPPTYSKSHGKTYGEHKHLEFSDQYRELQRYABEVGIFFTASGMDMAVE 120
Db 61 KFNKALERPPTYSKSHGKTYGEHKHLEFSDQYRELQRYABEVGIFFTASGMDMAVE 120
QY 121 FLHELNVPPFKVSGDGNPPYLEKTKKGRPMVSSGQSMQMDTMKVQVIVKPLNPNFC 180
Db 121 FLHELNVPPFKVSGDGNPPYLEKTKKGRPMVSSGQSMQMDTMKVQVIVKPLNPNFC 180
QY 181 FLQCTSAFLQPDVNLRLVISEYQKLFDPDIPIGYSGHETGTAISVAALGAKVLERHIT 240
Db 181 FLQCTSAFLQPDVNLRLVISEYQKLFDPDIPIGYSGHETGTAISVAALGAKVLERHIT 240
QY 241 LDKTWKGDHSALEPGEAELVRSVRLVERALGSPTKQLLPCEMACNEKLGKSVVAVK 300
Db 241 LDKTWKGDHSALEPGEAELVRSVRLVERALGSPTKQLLPCEMACNEKLGKSVVAVK 300
QY 301 IPEGTILTMDLTVKVGEPKAYPPEDIFNLVGGKVLVTVEEDDTIMEELVDNHGKKIKS 359
Db 301 IPEGTILTMDLTVKVGEPKAYPPEDIFNLVGGKVLVTVEEDDTIMEELVDNHGKKIKS 359

RESULT 8
AAM41772

ID AAM41772 standard; protein; 400 AA.

XX AC AAM41772;

XX DT 22-OCT-2001 (first entry)

XX DE Human polypeptide SEQ ID NO 6703.

XX KW Human; nontropic; immunosuppressant; cytostatic; gene therapy; cancer;
KW peripheral nervous system; neuropathy; central nervous system; CNS;
KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
KW leukaemia.

XX OS Homo sapiens.

XX PN WO200153312-A1.

XX PD 26-JUL-2001.

XX PF 26-DEC-2000; 2000WO-US034263.

XX PR 23-DEC-1999; 99US-00471275.

XX PR 21-JAN-2000; 2000US-00488725.

XX PR 25-APR-2000; 2000US-0052317.

XX PR 20-JUN-2000; 2000US-00598042.

XX PR 19-JUL-2000; 2000US-00620312.

XX PR 03-AUG-2000; 2000US-00653450.

XX PR 14-SEP-2000; 2000US-00662191.

XX PR 19-OCT-2000; 2000US-00693036.

XX PR 29-NOV-2000; 2000US-00727344.

XX PA (HYSE-) HYSEQ INC.

XX PI Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J, Zhao QA;
PI Zhou P, Goodrich R, Drmanac RT;

XX DR WPI; 2001-442253/47.

XX DR N-PSDB; AAI60928.

XX PT Novel nucleic acids and polypeptides, useful for treating disorders such
PT as central nervous system injuries.

XX PS Example 2; SEQ ID NO 6703; 10078pp; English.

XX CC The invention relates to human nucleic acids (AAI57798-AAI61369) and the
CC encoded polypeptides (AAM38642-AAM42213) with nontropic,
CC immunosuppressant and cytostatic activity. The polynucleotides are useful
CC in gene therapy. A composition containing a polypeptide or polynucleotide
CC of the invention may be used to treat diseases of the peripheral nervous
CC system, such as peripheral nervous injuries, peripheral neuropathy and
CC localised neuropathies and central nervous system diseases, such as
CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
CC utilisation of the activities such as: Immune system suppression,
CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
CC assays for receptor activity, arthritis and inflammation, leukaemias and
CC C.N.S disorders. Note: The sequence data for this patent did not form
CC part of the printed specification

XX SQ Sequence 400 AA;

Query Match 94.8%; Score 1789; DB 4; Length 400;

Best Local Similarity 88.3%; Pred. No. 1.2e-173;

Matches 346; Conservative 0; Mismatches 2; Indels 44; Gaps 1;

QY 1 MPELELCPRWVGQHPFCFIIAETGQNHQGDLDVAKRMIRMAKEGADCAKFKQSELEF 60

Db 9 MPELELCPRWVGQHPFCFIIAETGQNHQGDLDVAKRMIRMAKEGADCAKFKQSELEF 68

QY 61 KFNKALERPPTYSKSHGKTYGEHKHLEFSDQYRELQRYABEVGIFFTASGMDMAVE 120

Db 69 KFNKALERPPTYSKSHGKTYGEHKHLEFSDQYRELQRYABEVGIFFTASGMDMAVE 128

QY 121 FLHELNVPPFKVSGDGNPPYLEKTKKGRPMVSSGQSMQMDTMKVQVIVKPLNPNFC 148

Db 129 FLHELNVPPFKVSGDGNPPYLEKTKKGRPMVSSGQSMQMDTMKVQVIVKPLNPNFC 188

QY 149 -----KGRPMVSSGQSMQMDTMKVQVIVKPLNPNFCFLQCTSAFLQPDVNL 196

Db 189 KKVLMVLVDYSGRPMVSSGQSMQMDTMKVQVIVKPLNPNFCFLQCTSAFLQPDVNL 248

QY 197 LRVISEYQKLFDPDIPIGYSGHETGTAISVAALGAKVLERHITLDTWKGDHSALEP 256

Db 249 LRVISEYQKLFDPDIPIGYSGHETGTAISVAALGAKVLERHITLDTWKGDHSALEP 308

QY 257 GELAEIVRSVRLVERALGSPTKQLLPCEMACNEKLGKSVVAVKIPGTLTMDMLTVKV 316

Db 309 GELAEIVRSVRLVERALGSPTKQLLPCEMACNEKLGKSVVAVKIPGTLTMDMLTVKV 368

QY 317 GEPKAYPPEDIFNLVGGKVLVTVEEDDTIMEE 348

Db 369 GEPKAYPPEDIFNLVGGKVLVTVEEDDTIMEE 400

RESULT 9

ABBY71364

ID ABBY71364 standard; protein; 338 AA.

XX AC ABBY71364;

XX DT 26-MAR-2002 (first entry)

XX DE Drosophila melanogaster polypeptide SEQ ID NO 40884.

XX KW Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical.

XX OS Drosophila melanogaster.

XX PN WO200171042-A2.

XX PD 27-SEP-2001.

XX PF 23-MAR-2001; 2001WO-US009231.

XX XX

```
PR 23-MAR-2000; 2000US-0191637P.
PR 11-JUL-2000; 2000US-00614150.
XX (PEKE ) PE CORP NY.
PA Venter JC, Adams M, Li PWD, Myers EW;
XX
XX WPI; 2001-656860/75.
DR N-PSDB; ABL15467.
XX
XX New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signaling and cell-cell
PT interactions.
XX
XX Disclosure; SEQ ID NO 40884; 21pp + Sequence Listing; English.
PS
XX
XX The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABU16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins (ABBS7737-
CC ABBS72072). The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
XX SQ Sequence 338 AA;
Query Match 41.5%; Score 783; DB 4; Length 338;
Best Local Similarity 45.6%; Pred. No. 6.5e-71;
Matches 154; Conservative 60; Mismatches 100; Indels 24; Gaps 3;
QY 39 MIRMAKCGACAKFQKSELEFKNRALKALRPYTSKHSWGKTYGHEHLEHLEFSDQYREL 98
DB 1 MIWEAKGAGCHCVKFKSDDLPAKTRSDLDREYISDHAWGKTYGHEHLEFSDQYQL 60
QY 99 QRYAEVGIFFTAGMDMAVEFLHMLNVPFKVSGDNTNPPYLEKTAKKGRPMVSISSG 158
DB 61 QARCKELNVDTFASAMDRSLFSLNVPFKIGSGDANNFPLKKAANLNLPLVISTG 120
QY 159 MQSDMTKQVYQIVKPL-NPNFCLQCTSAYPELVNLRVISEYQKLFDPIDIGYSGH 217
DB 121 MNTQWTVVERIVQTVRESKEDYALMHCVSSYPTPKDCSLQISLVIRTRFPNVAIGYSGH 180
QY 218 ETGTAISVAAALGAKVLERHITLDTKWKGDHSGASLEPGLAEALVRSV----- 266
DB 181 ELGVIIISQAALLGARIVRHFLLDKSQGSDHRCSLFPQELKALTITAITNFKLSSVEMP 240
QY 267 --RLVERALGS-----PTKQLLPCEMACNEKLGKSVVAKVKIPGEGILTMDMLTV 314
DB 241 PQEIVKGLNGDEELEAALQHVESKTLPCFLPCRNKLGKSIVAARNLNKGYRLQADMAI 300
QY 315 KVGEPKAYPPEDIENLVGKKVLVTVEEDDTIMEELVDN 352
DB 301 KVSEPSGLTAEFDLVLGKELADNIGEDPEILNGSIIN 338
RESULT 10
AAV68965.
ID AAV68965 standard; protein; 338 AA.
XX
XX AAV68965;
XX
XX 30-MAY-2000 (first entry)
XX
XX Cps2P protein which is involved in stalic acid synthesis.
XX
XX Capsular gene cluster; serotype 2; polysaccharide biosynthesis;
KW capsular component; antigen; regulation; chain length determination;
KW complement-mediated opsonophagocytosis; serotype-specific detection;
KW antigen; vaccine; Streptococcal disease; ORF 2Z; ORF 2Y; ORF 2Z; Cps2A;
KW Cps2B; Cps2C; Cps2D; Cps2E; Cps2F; Cps2G; Cps2H; Cps2I; Cps2J; Cps2K;
```

```
XX Cps2O; Cps2P; Cps2Q; Cps2R; Cps2S; Cps2T.
XX Streptococcus suis.
XX WO200005378-A2.
XX 03-FEB-2000.
XX 19-JUL-1999; 99WO-NL000460.
XX 22-JUL-1998; 98EP-00202465.
XX 22-JUL-1998; 98EP-00202467.
XX (DIEN-) STICHTING DIENST LANDBOUWKUNDIG ONDERZOE.
XX Smith HE;
XX WPI; 2000-195104/17.
XX N-PSDB; AAZ60929.
XX New nucleic acid containing the capsular gene cluster of Streptococcus
PT suis, used for serotype-specific detection and to generate antigens or
PT mutants for vaccination.
XX
XX Disclosure; Fig 3; 144pp; English.
XX
XX The proteins AAY68950-69 are encoded by the capsular gene cluster of
CC Streptococcus suis serotype 2. The genes in this cluster are involved in
CC polysaccharide biosynthesis of capsular components and antigens. The
CC proteins are involved in regulation (CpsA), chain length determination
CC (CpsB, CpsC), export (CpsD), and biosynthesis (CpsE, CpsF, CpsG, CpsH,
CC CpsI, CpsK). The capsule confers bacterium resistance to complement-
CC mediated opsonophagocytosis. The gene cluster is used as a source of
CC probes and primers for serotype-specific detection of S. suis and is also
CC useful for recombinant production of the proteins. The proteins are then
CC useful for producing antigens that can be used in vaccines, for
CC controlling or eradicating a Streptococcal disease, in humans or animals,
CC e.g. against S. suis in pigs
XX
XX SQ Sequence 338 AA;
Query Match 28.9%; Score 544.5; DB 3; Length 338;
Best Local Similarity 36.0%; Pred. No. 1.6e-46;
Matches 123; Conservative 73; Mismatches 129; Indels 17; Gaps 8;
QY 20 FILAEIGNQHGDLDAKRMIRMAKCEGACAKFQKSELEFKNRALKALRPYTSKHSWGK 79
DB 3 YIIAEIGCNHNGDVHLARKMVEVAVDCGVDAVKFTFKADLLISKYAPKAEY-QKITTGE 61
QY 80 TYG--EHKHLHFSDQYRELQRYAEVGIFFTAGMDMAVEFLHMLNVPFKVSGDGT 137
DB 62 SDSQLEMTTRRLELSPFEEYLDLDYCLEKGVDFVSTFPDEESLDLFDLSTDMPVYKIPSEI 121
QY 138 NNFPYLEKTAKKGRPMVSISSGMSQSDMTKQVYQIVKPLNPN-----FCFLQCTSAYPEL 193
DB 122 TNLPLYLEKIGQAKVKVILSTGMAVND---EIHQAVKIIQENGTTDISILHCTTEYPTVP 178
QY 194 DVNLRVISEYQKLFDPIDIGYSGHETGTAISSVAAVALGAKVLERHITLDTKWKGDHSGAS 253
DB 179 ALNVLNLVTLTKKEFPNLITIGYSDHSGSEVPITAAAMGAELIEKHFTLIDNEMEGPDHKAS 238
QY 254 LEPGELAEVRSVRLVERALGSPKQLLPCEMACNEKL--GKSVVAKVKIPGEGILTMDM 311
DB 239 ATPDILAAALVKGVRIVEQSLGKFEK--PEEVEVRNKIVARKSIVAKIAKGEVFTTEEN 296
QY 312 LTVKVGEP-KAYPPEDIENLVGKKVLVTVEEDDTIMEELVDN 352
DB 297 ITVK--RPGNGISPMWYKVLGVQSEQDFEEDQNICHSAFEN 336
RESULT 11
ABP26810
ID ABP26810 standard; protein; 341 AA.
```


XX SQ Sequence 346 AA;

Query Match 27.1%; Score 510.5; DB 3; Length 346;

Best Local Similarity 36.3%; Pred. No. 4.9e-43;

Matches 123; Conservative 61; Mismatches 140; Indels 15; Gaps 6;

QY 20 FIIAIEGQNHQDLDVAKRMIRMAKCGADCAKFKQSELEFFKFNKALERTPYTSKHSWG- 78

DB 5 YIVAEIGCNHNGSVDIAREMILKAKEAGVNAVVFQTFKADKLISAIAPKAEYQIKNTGEL 64

QY 79 KTYGEHRHLEFSDHQRELQRYAEVGIFFTAGMDMAVEFELHNLVFFKVGSGDTN 138

DB 65 ESQLEMTKLEMYDDYHLHMEYAVSLNLDVFFDFDESDIDFLASLKQIKWIPSGELL 124

QY 139 NFPVLEKTAKKGRP---WVSSGQSMQMDTMKVQVQIV---KPLNPNFCFLQCTSAYPELP 192

DB 125 NLPVLEKIAKLPIPDKKIIISTGMATIDEIKQSVSIFINNKPVGNTILHCNTEYPTPF 184

QY 193 EDVNLRLVISEYQKLPDPIPIGYSGHETGIAISVAALGAKVLERHITLDTKWGSDHSA 252

DB 185 EDVNLNAINDLKHPKNNIGFSDHSSGFYAAIAAVPYGITFIEKHFTLDKSMSPDHLA 244

QY 253 SLEPGEIAELVRSVRLVERALGSPKQLPCMACNEKLGKSVVAKVKIPSGTILTMDEL 312

DB 245 SIEPDELKHLGICVRCVEKSLGNSKVVVTASERKNKIVARKSIIAKTEIKKGEVFSEKNI 304

QY 313 TVKVGEP-KAYPPEDIFNLVGKVLVTVEEDDTIMEELV 350

DB 305 TTK--RFGNGISPMEWYNLLGK-----IAEQDFIPDELI 336

RESULT 13

AAB84684

ID AAB84684 standard; protein; 346 AA.

AC AAB84684;

XX 17-SEP-2001 (first entry)

XX Amino acid sequence of a bacterial sialic acid synthetase.

XX Carbohydrate processing; sialylation; sialylated glycoprotein; CMP-SA;

XX cytidine monophosphate-sialic acid; N-acetylglucosamine-2 epimerase;

XX sialic acid synthetase; aldolase; CMP-SA synthetase; CMP-SA transporter;

XX vaccine.

XX Escherichia coli.

XX WO200142492-A1.

XX 14-JUN-2001.

XX 07-DEC-2000; 2000WO-US033136.

XX 09-DEC-1999; 99US-0169839P.

XX (HUMA-) HUMAN GENOME SCI INC.

XX (UYJO) UNIV JOHNS HOPKINS.

XX (UTEM) UNIV TEMPLE.

XX (UYWY-) UNIV WYOMING.

XX Betenbaugh MJ, Lawrence S, Lee YC, Coleman TA, Palter K;

XX Jarvis D;

XX WPI; 2001-441575/47.

XX DR N-PSDB; AAH28459.

XX Cells producing cytidine monophosphate-sialic acid and sialylated

XX glycoprotein above endogenous levels for production of vaccines and

XX therapeutics.

XX Example 5; Page 165-166; 182pp; English.

XX The specification describes a method for manipulating carbohydrate

XX processing pathways in cells of interest. The methods are used to

XX manipulate multiple pathways involved with the sialylation reaction by

XX using recombinant DNA technology and substrate feeding approaches to

XX enable the production of sialylated glycoproteins in the cells. The

XX sialylation process involves the post-translational addition of the donor

XX substrate cytidine monophosphate-sialic acid (CMP-SA) onto a specific

XX acceptor carbohydrate. The cells express at least one enzyme, selected

XX from N-acetylglucosamine-2 epimerase, sialic acid synthetase, aldolase,

XX CMP-SA synthetase and CMP-SA transporter, above endogenous levels. The

XX cells are useful for producing complex sialylated glycoproteins in cells

XX of interest, especially insect cells. Glycoproteins containing sialylated

XX oligosaccharides are useful as vaccines, therapeutics and diagnostic

XX tools. Cells producing complex sialylated glycoproteins are useful for

XX enhancing the value of heterologous expression systems and increasing the

XX application of heterologous cell expression products as vaccines,

XX therapeutics and diagnostic tools as well as increasing the variety of

XX heterologous proteins that can be produced and lowering biotechnology

XX production costs. The present sequence represents a sialic acid

XX synthetase (neub), which is used in the method of the invention

XX SQ Sequence 346 AA;

Query Match 27.1%; Score 510.5; DB 4; Length 346;

Best Local Similarity 36.3%; Pred. No. 4.9e-43;

Matches 123; Conservative 61; Mismatches 140; Indels 15; Gaps 6;

QY 20 FIIAIEGQNHQDLDVAKRMIRMAKCGADCAKFKQSELEFFKFNKALERTPYTSKHSWG- 78

DB 5 YIVAEIGCNHNGSVDIAREMILKAKEAGVNAVVFQTFKADKLISAIAPKAEYQIKNTGEL 64

QY 79 KTYGEHRHLEFSDHQRELQRYAEVGIFFTAGMDMAVEFELHNLVFFKVGSGDTN 138

DB 65 ESQLEMTKLEMYDDYHLHMEYAVSLNLDVFFDFDESDIDFLASLKQIKWIPSGELL 124

QY 139 NFPVLEKTAKKGRP---WVSSGQSMQMDTMKVQVQIV---KPLNPNFCFLQCTSAYPELP 192

DB 125 NLPVLEKIAKLPIPDKKIIISTGMATIDEIKQSVSIFINNKPVGNTILHCNTEYPTPF 184

QY 193 EDVNLRLVISEYQKLPDPIPIGYSGHETGIAISVAALGAKVLERHITLDTKWGSDHSA 252

DB 185 EDVNLNAINDLKHPKNNIGFSDHSSGFYAAIAAVPYGITFIEKHFTLDKSMSPDHLA 244

QY 253 SLEPGEIAELVRSVRLVERALGSPKQLPCMACNEKLGKSVVAKVKIPSGTILTMDEL 312

DB 245 SIEPDELKHLGICVRCVEKSLGNSKVVVTASERKNKIVARKSIIAKTEIKKGEVFSEKNI 304

QY 313 TVKVGEP-KAYPPEDIFNLVGKVLVTVEEDDTIMEELV 350

DB 305 TTK--RFGNGISPMEWYNLLGK-----IAEQDFIPDELI 336

RESULT 14

AAO26547

ID AAO26547 standard; protein; 346 AA.

XX AAO26547;

XX 06-MAR-2003 (first entry)

XX Bacterial sialic acid synthetase (NeuB) protein.

XX Vaccine; glycoprotein; insect cell; enzyme; N-acetylglucosamine-2;

XX GlcNAc-2; epimerase; UDP-GlcNAc; mannose; (Man)NAC; sialic acid;

XX synthetase; aldolase; cytidine monophosphate-sialic acid; CMP-SA;

XX transporter; sialylated glycoprotein.

XX Escherichia coli.

XX US2002142386-A1.

XX 03-OCT-2002.

```
XX PF 16-AUG-2001; 2001US-00930440.
XX PR 02-MAR-1999; 99US-0122582P.
XX PR 08-DEC-1999; 99US-0169624P.
XX PR 25-AUG-2000; 2000US-0227579P.
XX PA (BETE/) BETENBAUGH M J.
XX PA (LAWR/) LAWRENCE S.
XX PA (LEEV/) LEE Y C.
XX PA (COLE/) COLEMAN T A.
XX PI Betenbaugh MJ, Lawrence S, Lee YC, Coleman TA;
XX DR WPI; 2003-102519/09.
XX DR N-PSDB; AAL53994.
XX PT Manipulating glycoprotein production in insect cell, involves enhancing
XX PT expression of enzymes involved in carbohydrate processing pathway such as
XX PT N-acetylglucosamine-2 epimerase or sialic acid synthetase.
XX PS Disclosure; Fig 35D; 88pp; English.
XX CC The invention relates to a novel method for manipulating glycoprotein
XX CC production in an insect cell comprising enhancing expression of an
XX CC enzyme, such as N-acetylglucosamine-2 (GlcNAc-2) epimerase, one
XX CC catalysing conversion of UDP-GlcNAc to mannose (Man)NAC, sialic acid
XX CC synthetase, aldolase, cytidine monophosphate-sialic acid (CMP-SA)
XX CC synthetase or CMP-SA transporter, where the expression of each enzyme is
XX CC enhanced to above endogenous levels. The novel method is useful for
XX CC manipulating glycoprotein production in an insect cell. Further methods
XX CC of the invention are useful for producing sialylated glycoprotein. The
XX CC sialylated glycoprotein produced by the above mentioned methods are
XX CC useful as pharmaceutical compositions, vaccines, diagnostics and
XX CC therapeutics. This sequence represents the bacterial sialic acid (NeuB)
XX CC synthetase protein of the invention
XX SQ Sequence 346 AA;
Query Match 27.1%; Score 510.5; DB 6; Length 346;
Best Local Similarity 36.3%; Pred. No. 4.9e-43;
Matches 123; Conservative 61; Mismatches 140; Indels 15; Gaps 6;
QY 20 FTIABIGNHQGLDVAKRMIRMAKECGADCAKFKSELEFKFNKALERPYSKHSWG- 78
DB 5 YIVABIGCNHNGSVDIAREMILKAKEAGVNAVVFQTFKADKLISAIAPAKAEYQIKNTGEL 64
QY 79 KTYGKHKLHLEFSDQYRELQRYAEVGIFFTASGDMEMAVFLHNLNVPFFKVGSGDTN 138
DB 65 ESQLEWTKKLEMYDDYLHLMYAVSLNLDVFPDSDSIDFLASLKQKIWKIPSGELL 124
QY 139 NPPYLEKTAKKGRP---MVISGMSQMDTMKQVQIV---KPLNPNFCFLOCTSAVPLQP 192
DB 125 NLPLYELKIAKLPIPKKIIISTGMATIDIEIKQSVSIFINNKVPVGNITILHCNTYPTPPF 184
QY 193 EDVNLRVISEYQKLPDPDIPIGVSGHETGIAISVAVALGAKYLERHITLDTKWGSDHSA 252
DB 185 EDVNLNAINDLKHKHFPKNNIGFSDHSGFYAIAAAPPYGITIEKHFTLDKSMGPDHLA 244
QY 253 SLEPELAELVRSVRLVERALGSPTKQLLPCEMACNEKLGKSVAKVIPGEGTILTMOML 312
DB 245 SIEDELKHLICIGVRCVSEKSGNSKVTVASERKNKIVARKSIIATIKKGEVFSKNI 304
QY 313 TKVKGEP-KAYPPBEDIFNLVGKVLVTVEEDDTIMEELV 350
DB 305 TTK--RPGNGISPMWYNLLGK-----IAEQDFIPDELI 336
RESULT 15
ABG17251
ID ABG17251 standard; protein; 125 AA.
XX AC ABG17251;
```

```
XX DT 18-FEB-2002 (first entry)
XX DE Novel human diagnostic protein #17242.
XX KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
XX KW food supplement; medical imaging; diagnostic; genetic disorder.
XX OS Homo sapiens.
XX PN WO200175067-A2.
XX PD 11-OCT-2001.
XX PF 30-MAR-2001; 2001WO-US008631.
XX PR 31-MAR-2000; 2000US-00540217.
XX PR 23-AUG-2000; 2000US-00649167.
XX PA (HYSE-) HYSEQ INC.
XX PI Drmanac RT, Liu C, Tang YT;
XX DR WPI; 2001-639362/73.
XX DR N-PSDB; AAS81438.
XX PT New isolated polynucleotide and encoded polypeptides, useful in
XX PT diagnostics, forensic, gene mapping, identification of mutations
XX PT responsible for genetic disorders or other traits and to assess
XX PT biodiversity.
XX PS Claim 20; SEQ ID NO 47610; 103pp; English.
XX CC The invention relates to isolated polynucleotide (I) and polypeptide (II)
XX CC sequences. (I) is useful as hybridisation probes, polymerase chain
XX CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
XX CC and in recombinant production of (II). The polynucleotides are also used
XX CC in diagnostics as expressed sequence tags for identifying expressed
XX CC genes. (I) is useful in gene therapy techniques to restore normal
XX CC activity of (II) or to treat disease states involving (II). (II) is
XX CC useful for generating antibodies against it, detecting or quantitating a
XX CC polypeptide in tissue, as molecular weight markers and as a food
XX CC supplement. (II) and its binding partners are useful in medical imaging
XX CC of sites expressing (II). (I) and (II) are useful for treating disorders
XX CC involving aberrant protein expression or biological activity. The
XX CC polypeptide and polynucleotide sequences have application of mutations
XX CC diagnostics, forensic, gene mapping, identification of mutations
XX CC responsible for genetic disorders or other traits to assess biodiversity
XX CC and to produce other types of data and products dependent on DNA and
XX CC amino acid sequences. ABG00010-ABG30377 represent novel human diagnostic
XX CC amino acid sequences of the invention. Note: The sequence data for this
XX CC patent did not appear in the printed specification, but was obtained in
XX CC electronic format directly from WIPO at
XX CC ftp.wipo.int/pub/published_pct_sequences
XX SQ Sequence 125 AA;
Query Match 22.2%; Score 419.5; DB 4; Length 125;
Best Local Similarity 79.2%; Pred. No. 2e-34;
Matches 80; Conservative 6; Mismatches 14; Indels 1; Gaps 1;
QY 1 MPLELELCFGRWVGQHPFCFIIAEIGQNHQGLDVAKRMIRMAKECGADCAKFKSELEF 60
DB 9 MPVELELCFGRWVGQHPFCFIIAEIGQNHQGLTVAKRMIRMAKECGADCAKFKSELEF 68
QY 61 KFNKALERPYSKHSWGK-TYGEHKHLEFSDQYRELQ 100
DB 69 KFNKALDRPYSKHSGLGEDVRGAQTDIWKFNHHDVQGAAR 109
Search completed: September 13, 2004, 14:41:01
Job time : 130 secs
```

GenCore version 5.1.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 13, 2004, 14:34:23 ; Search time 118 Seconds
(without alignments)
959.924 Million cell updates/sec

Title: US-10-759-277-4

Perfect score: 1897

Sequence: 1 MPLELELCPRWVGQHPHF.....BEDDTIMEELVDNHHKKIKS 359

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL 25:*

- 1: sp_archea:*
- 2: sp_bacteria:*
- 3: sp_fungi:*
- 4: sp_human:*
- 5: sp_invertebrate:*
- 6: sp_mammal:*
- 7: sp_mbc:*
- 8: sp_organelle:*
- 9: sp_phase:*
- 10: sp_plant:*
- 11: sp_rodent:*
- 12: sp_virus:*
- 13: sp_vertebrate:*
- 14: sp_unclassified:*
- 15: sp_rvirus:*
- 16: sp_bacteriap:*
- 17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1810	95.9	359	11 Q99J77	Q99J77 mus musculus
2	1804	95.6	359	11 Q99JH0	Q99JH0 mus musculus
3	854	45.3	372	5 Q9VG74	Q9VG74 chrosophila
4	547.5	29.0	280	16 Q8KE65	Q8KE65 chlorobium
5	534.5	28.3	341	2 Q87186	Q87186 streptococc
6	534.5	28.3	341	16 Q8E501	Q8E501 streptococc
7	532.5	28.2	339	2 Q9ALW6	Q9ALW6 streptococc
8	532.5	28.2	339	2 Q93TI2	Q93TI2 streptococc
9	532.5	28.2	341	2 Q9RPC0	Q9RPC0 streptococc
10	532.5	28.2	341	16 Q8DZ33	Q8DZ33 streptococc
11	526	27.9	749	16 Q82UC3	Q82UC3 nitrosomona
12	524	27.8	311	16 Q82HY4	Q82HY4 streptomyc
13	518	27.5	334	2 Q7X523	Q7X523 campylobact
14	516.5	27.4	333	16 Q8DD27	Q8DD27 vibrio vuln
15	514	27.2	361	2 Q8KNA2	Q8KNA2 pseudomonas
16	512	27.1	312	16 Q9AK45	Q9AK45 streptomyc

ALIGNMENTS

RESULT 1

Q99J77 PRELIMINARY; PRT; 359 AA.

AC Q99J77; TREMBLrel. 17, Created

DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)

DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)

DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)

DE Similar to N-acetylneuraminic acid phosphate synthase, sialic acid synthase.

DE GN NANS OR 4632418E04RIK OR SAS.

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.

OC NCBI_TaxID=10090;

OX [1]_TaxID=10090;

RN [1] SEQUENCE FROM N.A.

RP SEQUENCE FROM N.A.

RA Strausberg R.;

RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.

RN [2] SEQUENCE FROM N.A.

RC STRAIN=C57BL/6J; TISSUE=Skin;

RX MEDLINE=22354683; PubMed=1246851;

RA The FANTOM Consortium,

RA "Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs."

RT Nature 420:563-573 (2002).

RL Nature 420:563-573 (2002).

DR EMBL; BC003307; AAH03307.1; -.

DR EMBL; AK076290; BAC36290.1; -.

DR HSSP; P19614; 2J1A.

DR MGD; MGI:2149820; Nans.

DR GO; GO:0005829; C:cytosol; IDA.

DR GO; GO:0019007; F:N-acetylneuraminic acid phosphate synthase . . . ; IDA.

DR InterPro; IPR006013; AntifreezeIII.

DR InterPro; IPR006014; Antifreeze-Gom.

DR InterPro; IPR006190; Antifreeze_like.

DR InterPro; IPR004144; NeuB.

DR Pfam; PF01354; Antifreeze; 1.

DR Pfam; PF03102; NeuB; 1.

Q46675 escherichia

Q9rdx5 legionella

Q893u6 clostridium

Q8f5q4 leptospira

Q9aig6 streptococ

Q89hl9 bradyrhizob

Q7u964 synechococc

Q9pmx2 campylobact

Q7u911 synechococc

Q93nq2 escherichia

Q89hj8 bradyrhizob

Q8tjl5 methanosarc

Q8f5q8 leptospira

Q8a7l1 bacteroides

Q87t70 vibrio para

Q87h26 clostridium

Q8kh52 pseudomonas

Q8r9s2 aeromonas p

Q8rib8 fusbacteri

Q8kn54 pseudomonas

Q57265 neisseria m

Q8f330 leptospira

Q7v953 prochloroco

Q939j8 campylobact

Q9r9m1 rhizobium m

Q9f9f4 campylobact

Q9a4g6 caulobacter

Q9pmv2 campylobact

Q9eu02 campylobact

[illegible]

Qy 249 DHSASLEPGEAELVRSVRVLVERALGSPTKQLLPCEMACNEKL 291
 - : :: - :
D6 235 DQAASVEISGMSRSLVSNIRDIEKALGDGVKRVDGEAAARKKL 277

RESULT 5

087186 PRELIMINARY; PRT; 341 AA.
AC 087186;
AC 087186; (TRENBLrel. 08, Created)
DT 01-NOV-1998 (TRENBLrel. 08, Last sequence update)
DT 01-JUN-2003 (TRENBLrel. 24, Last annotation update)
DE NEUB protein.
DE GN NEUB.
OS Streptococcus agalactiae.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1311;
RN [1]
RP SEQUENCE FROM N.A.
RA Yamamoto S., Miyake K., Iijima S.;
RT Identification and Characterization of cps (capsular po
RT Genes from Streptococcus agalactiae type Ia.;
RL Submitted (SEP-1998) to the EMBL/GenBank/DDAY databases.

RP SEQUENCE FROM N. A.

RA Miyake K., Yamamoto S., Koike Y., Watanabe M., Iijima S.;
 RT "Molecular Characterization of Type-Specific Capsular Polysaccharide
 RT Biosynthesis Genes of *Streptococcus agalactiae* Type Ia.";
 RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.

DR GO: 0016051: P: carbohydrate biosynthesis: TEA

DR InterPro: IPR006014: Antifreeze dom:
DR CO:0010051, F: carbonylate biosyn

DR InterPro; IPR006190; Antifreeze like.

DR InterPro; IPR004144; NeuB.

DR Pfam; PF01354; Antifreeze; 1.

DR pfam; PF03102; NeuB; 1.

```
DR PROSITE; PS50844; APP_LIKE; 1
SQ SEQUENCE 241 AA. 38091 MW.
```

SQ SEQUENCE 341 AA; 38091 MW;

Query Match 28.38:

Best Local Similarity 36.8%

Matches 125; Conservative 66

100

QY 20 FIIAEIGQNHQGLDVAKRM

[illegible]

DB 3 Y I A E I G C N H N G D I N L A R K M

Qy 80 TYG--EHKRHLEFSHDOYRE

0
1
2
3
4
5
6
7
8
9
A
B
C
D
E
F
G
H
I
J
K
L
M
N
O
P
Q
R
S
T
U
V
W
X
Y
Z

Db 62 ADSQLEMTKRLELSFEEYLE

100

QY
138 NNFPYLEKTAKKGRPMVSS

rk
122 TN1 PVI FYIC KCKKVVII STC

DB IZZ INLFYLEXIGKQKRVILSTC

Ov 194 DVNLRVISEYOKLFPDIPIGY

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
84

Db 179 SLNLNVIHTLKDEFKDLTIGV

100

Qy 254 LEPGELAE LVR SVRLVERALC

239 ATPDILAAI VKGVRIVECALC

DB 239 AIPDILAAALVKGVRIVEQALN

QY 314 VKVGEP-KAYPPEDIENLVGI

[illegible]

Db 299 VK--RPGNGISPMNWDILG

1
 2
 3
 4
 5
 6
 7
 8
 9
 10
 11
 12
 13
 14
 15
 16
 17
 18
 19
 20
 21
 22
 23
 24
 25
 26
 27
 28
 29
 30
 31
 32
 33
 34
 35
 36
 37
 38
 39
 40
 41
 42
 43
 44
 45
 46
 47
 48
 49
 50
 51
 52
 53
 54
 55
 56
 57
 58
 59
 60
 61
 62
 63
 64
 65
 66
 67
 68
 69
 70
 71
 72
 73
 74
 75
 76
 77
 78
 79
 80
 81
 82
 83
 84
 85
 86
 87
 88
 89
 90
 91
 92
 93
 94
 95
 96
 97
 98
 99
 100
 101
 102
 103
 104
 105
 106
 107
 108
 109
 110
 111
 112
 113
 114
 115
 116
 117
 118
 119
 120
 121
 122
 123
 124
 125
 126
 127
 128
 129
 130
 131
 132
 133
 134
 135
 136
 137
 138
 139
 140
 141
 142
 143
 144
 145
 146
 147
 148
 149
 150
 151
 152
 153
 154
 155
 156
 157
 158
 159
 160
 161
 162
 163
 164
 165
 166
 167
 168
 169
 170
 171
 172
 173
 174
 175
 176
 177
 178
 179
 180
 181
 182
 183
 184
 185
 186
 187
 188
 189
 190
 191
 192
 193
 194
 195
 196
 197
 198
 199
 200
 201
 202
 203
 204
 205
 206
 207
 208
 209
 210
 211
 212
 213
 214
 215
 216
 217
 218
 219
 220
 221
 222
 223
 224
 225
 226
 227
 228
 229
 230
 231
 232
 233
 234
 235
 236
 237
 238
 239
 240
 241
 242
 243
 244
 245
 246
 247
 248
 249
 250
 251
 252
 253
 254
 255
 256
 257
 258
 259
 260
 261
 262
 263
 264
 265
 266
 267
 268
 269
 270
 271
 272
 273
 274
 275
 276
 277
 278
 279
 280
 281
 282
 283
 284
 285
 286
 287
 288
 289
 290
 291
 292
 293
 294
 295
 296
 297
 298
 299
 300
 301
 302
 303
 304
 305
 306
 307
 308
 309
 310
 311
 312
 313
 314
 315
 316
 317
 318
 319
 320
 321
 322
 323
 324
 325
 326
 327
 328
 329
 330
 331
 332
 333
 334
 335
 336
 337
 338
 339
 340
 341
 342
 343
 344
 345
 346
 347
 348
 349
 350
 351
 352
 353
 354
 355
 356
 357
 358
 359
 360
 361
 362
 363
 364
 365
 366
 367
 368
 369
 370
 371
 372
 373
 374
 375
 376
 377
 378
 379
 380
 381
 382
 383
 384
 385
 386
 387
 388
 389
 390
 391
 392
 393
 394
 395
 396
 397
 398
 399
 400
 401
 402
 403
 404
 405
 406
 407
 408
 409
 410
 411
 412
 413
 414
 415
 416
 417
 418
 419
 420
 421
 422
 423
 424
 425
 426
 427
 428
 429
 430
 431
 432
 433
 434
 435
 436
 437
 438
 439
 440
 441
 442
 443
 444
 445
 446
 447
 448
 449
 450
 451
 452
 453
 454
 455
 456
 457
 458
 459
 460
 461
 462
 463
 464
 465
 466
 467
 468
 469
 470
 471
 472
 473
 474
 475
 476
 477
 478
 479
 480
 481
 482
 483
 484
 485
 486
 487
 488
 489
 490
 491
 492
 493
 494
 495
 496
 497
 498
 499
 500
 501
 502
 503
 504
 505
 506
 507
 508
 509
 510
 511
 512
 513
 514
 515
 516
 517
 518
 519
 520
 521
 522
 523
 524
 525

RESULT 6
C9FE01

Q8E301 ;

ID	Q8E501	PRELIMINARY;	PRT;	341 AA.
AC	Q8E501;			
DT	01-MAR-2003	(TREMBLrel. 23, Created)		
DT	01-MAR-2003	(TREMBLrel. 23, Last sequence update)		
DT	01-JUN-2003	(TREMBLrel. 24, Last annotation update)		
DE	Hypothetical protein.			
GN	NEUB OR GRS1236.			
OS	Streptococcus agalactiae (serotype III).			
OC	Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;			
OC	Streptococcus.			
OX	NCBI_TaxID=216495;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN-NEM315 / Serotype III;			
RX	MEDLINE=22242508; PubMed=12354221;			
RA	Glaser P., Rusnlok C., Buchrieser C., Chevalier F., Frangeul L.,			
RA	Msadek T., Zouine M., Couve E., Lalloui L., Poyart C., Trieu-Cuot P.,			
RA	Kunst F.;			
RT	Genome sequence of Streptococcus agalactiae, a pathogen causing			
RT	invasive neonatal disease.;			
RL	Mol. Microbiol. 45:1499-1513 (2002).			
DR	EMBL; AL766849; CAD46895.1; -.			
DR	SAGAList; gbs1236; -.			
DR	GO; GO:0016051; P:carbohydrate biosynthesis; IEA.			
DR	InterPro; IPR006014; Antifreeze_dom.			
DR	InterPro; IPR006190; Antifreeze_like.			
DR	InterPro; IPR004144; NeUB.			
DR	Pfam; PF013354; Antifreeze; 1.			
DR	Pfam; PF03102; NeUB; 1.			
DR	PROSITE; PS50844; AFP_LIKE; 1.			
KW	Hypothetical protein; Complete proteome.			
SQ	SEQUENCE 341 AA; 38091 MW; 17620BF82A2EA340 CRC64;			
Query Match 28.3%; Score 534.5; DB 16; Length 341;				
Best Local Similarity 36.8%; Pred. No. 1.1e-35; Indels 13; Gaps 6				
Matches 125; Conservative 66; Mismatches 136;				
Qy	20	FIIAEIGNHOGDGLDVAKRMIRMAKECGADCAKFKQSELEFKFNKALERPYTSKHSWGK	79	
Db	3	YIIAEIGNHNGDINLAKMWVDVAVSCGVDAVKQTPKAEKLISKFAPKAEY-QKETGT	61	
Qy	80	TYG--EKKRHLEFSDQYRELQRYAEVGIETFTAGMDMAVEFLHNLVFFKVGSGDT	137	
Db	62	ADSQLEMTKRLSELFEEYLEMRDYAIKSGVETFTPTPEDESELEFLISTDMPYIKTSPGEI	121	
Qy	138	NNFYLEKTAKKGRPMWISSQMQMDTWKQYQVVKLPNP-FCFLQCTSAYPLOPE	193	
Db	122	TNLYLEKIKGQOKVILSTGMA--VMEEIHQAVNLRQNGTTDISLHCTTEYTPYP	178	
Qy	194	DVNLURVISEYQKLPDPIDPIGVSGHETGTIAISVAVALGAKVLERHITLDTKWGSDHSAS	253	
Db	179	SLNLNVHTLTKDFEKDTLTIGYSDHSIGSEVPIAAAMGAEVIEKHFTLDTNMEGPDHKAS	238	
Qy	254	LEPGELAEVRSVELVERALGSPTKQLLPCEMACNEKLGKSVAKVKIPEGTILTMDLT	313	
Db	239	ATPDILAALVKGVRIVQALGRFEPDPVBEKNKIVARKSVVAKPIKKGDIVSIENIT	298	
Qy	314	VKGEP-KAYPPEDIFNLVGKKVLTVVEEDDTIMEELVDN	352	
Db	299	VK--RPGNGISPMWYDILGQEAQDDEEDVIRDSRFEN	336	
RESULT 7				
Q9ALW6				
ID	Q9ALW6	PRELIMINARY;	PRT;	339 AA.
AC	Q9ALW6;			
DT	01-JUN-2001	(TREMBLrel. 17, Created)		
DT	01-JUN-2001	(TREMBLrel. 17, Last sequence update)		
DT	01-JUN-2003	(TREMBLrel. 24, Last annotation update)		
DE	Putative N-acetyl neuraminic acid synthetase NeUB.			
GN	NEUB.			
OS	Streptococcus agalactiae.			
OC	Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;			

```

OC Streptococcus.
OX NCBI_TaxID=1311;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NT6;
RA McKinnon K., Chaffin D.O., Rubens C.E.;
RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF337958; AAK11670.1; --. biosynthesis; IEA.
DR GO; GO:0016051; P:carbohydrate biosynthesis; IEA.
DR InterPro; IPR006014; Antifreeze dom.
DR InterPro; IPR006190; Antifreeze like.
DR Pfam; PF01354; Antifreeze; 1.
DR Pfam; PF03102; Antifreeze; 1.
DR PROSITE; PS0844; AFP LIKE; 1.
SQ SEQUENCE 339 AA; 37791 MW; 156C2DEBF1D453C7 CRC64;

Query Match      28.2%; Score 532.5; DB 2; Length 339;
Best Local Similarity 36.8%; Pred. No. 1.6e-35;
Matches 125; Conservative 66; Mismatches 136; Indels 13; Gaps 6;

QY 20 FTIIEIGQNHGDLVAKEMIRMAKCEGADCAKFKSELEFKNRKALRPPYTSKHSWGK 79
Db 3 YIIAIEGCHNGDINLAKQWVAVSCGVDAVKFQTFKAEKLSKFAKAEY-QKATTGT 61
QY 80 TYG--EHKHLEFSDHQRELYAEVEGIFTASGMDMAVEFLHNLVPPFKVSGGDT 137
Db 62 ADSQLEMTKRLSEFEEYLEMDRYAISKGVETFTFPDESELEFLISTDMPYIKPSGEI 121
QY 138 NNFPYLEKTAKGRPMWISSGMSQMDTKQVQIVKPLNPN---FCFLOCTSAYPELQPE 193
Db 122 TNLPYLEKTAKGRPMWISSGMSQMDTKQVQIVKPLNPN---FCFLOCTSAYPELQPE 178
QY 194 DYNLAVISEYQKLPDIPIGYSGHETGIAISVAALGAKVLERHITLDTKWKSGDHSAS 253
Db 179 SUNLNVHTLKDEKDLTIGYSDHSIGSEVPAAAAAGAEVIEKHFTLDTNMEGPDHKAS 238
QY 254 LEPGLAEIVRSVRLVERALGSPTKQLLPCENACNEKLGKSVVAKVPIEGTILTMDMIT 313
Db 239 ATPDILAAVKGRIVEQALGRFEKIPDVEEKNKIVARKSVVAKPKIKKGDYISIENT 298
QY 314 VKVGP-KAYPEDIFNLVKGKVLTVREDDTIMBELVDN 352
Db 299 VK--RPGNGISPMWYDILGQEAQDDFEDEVRDSRFEN 336

RESULT 9
QYRQPCO PRELIMINARY; PRT; 341 AA.
AC Q9RQPCO;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-JUN-2003 (TREMBLrel. 24, Last annotation update)
DE CPSN (NEUB).
GN CPSN OR NEUB.
OS Streptococcus agalactiae.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1311;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=COH1;
RX MEDLINE=93360815; PubMed=8355611;
RA Rubens C.E., Heggen L.M., Haft R.F., Wesels M.R.;
RT "Identification of cpsD, a gene essential for type III capsule expression in group B streptococci.";
RL Mol. Microbiol. 8:843-855(1993).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=COH1;
RA Chaffin D.O., Yim H.H., Beres S.B., Sweet E.S., Nittayajarn A.,
RA Rubens C.E.;
RT "Capsular Polysaccharide Synthesis Locus of Streptococcus agalactiae, Serotype III.";
RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=CNCTC 1/82;
RA McKinnon K., Chaffin D.O., Rubens C.E.;
RT "Streptococcus agalactiae type V polysaccharide synthesis operon complete sequence.";
RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF163833; AAD53074.1; --.
DR EMBL; AF349539; AAK29661.1; --.
DR PIR; T4651; T4651.
DR GO; GO:0016051; P:carbohydrate biosynthesis; IEA.
DR InterPro; IPR006014; Antifreeze dom.
DR InterPro; IPR006190; Antifreeze like.
DR InterPro; IPR004144; NeuB.

```

```

DR Pfam; PF01354; Antifreeze; 1.
DR Pfam; PF03102; NeuB; 1.
DR DR PROSITE; PS00844; APP_LIKE; 1.
SQ SEQUENCE 341 AA; 38033 MW; 148B756C2B2BF314 CRC64;

Query Match      28.2%; Score 532.5; DB 2; Length 341;
Best Local Similarity 36.8%; Pred. No. 1.6e-35;
Matches 125; Conservative 66; Mismatches 136; Indels 13; Gaps 6

Qy 20 FIIAIEIGNHGGDDLVAKRMTRMAKECGACADCAKQKSELEKFKRNKALERYPTSKHSGWK 79
Db 3 YIIAIEIGNHNGDINLAKKMYDVVASCVGDAVKQTGFKAEKLIISFPAPKABY-QKATGTG 61
Qy 80 TYG--EHRKHLEFSDHQVRELQRYAEVGIFFTAGMDMAVEFHLNLNVPFFKVGSGDT 137
Db 62 ADSQLEMTKRLSELFEEYLEMRDYAISKGVETFFSPFDEESLEFLISTDMPIYKPSGEI 121
Qy 138 NNFYLEKTAAGKGPWVISSGQMSQMDTKQVQIVKPLNP-----FCFLQCTSAYPLOQE 193
Db 122 TNLPLYEKIGHGQKKVILISTGMA--VNEETHQAVNIIURQNGTTDISILHCTTEYFTPTYP 178
Qy 194 DVNLRIVSEYOKLPFDPIDIGYSGHETGTAISVAVALGAKVLERHITLDTKWGSDHSA 253
Db 179 SLNLNVLIHTLKDEFKDLTGIGYSDHSIGSEVPITAAAMGAETVKHFTLDTNWEGPDHKAS 238
Qy 254 LEPCELAELVSRVLVERALGSPFKQLLPCEMACNEKLGKSVKAVKIPGCTILTMDLT 313
Db 239 ATPDILAAALVAGVRIVEQALGRFEKIPDPVEKKKIVARKSVALKPIKKGDIYSIENIT 298
Qy 314 VKVGEP-KAYPPEDIFNLVGGKVLVVEEDDTIMEELVDN 352
Db 299 VK--RPGNGISPMNWDILGEOAODDFEEDVIRDSREEN 336

```

RESULT 10	
Q8DZE3	
ID	PRELIMINARY; PRT; 341 AA.
Q8DZE3	
AC	Q8DZE3;
DT	01-MAR-2003 (TrEMBLrel. 23, Created)
DT	01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT	01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE	N-acetyl neuramic acid synthetase NeuB.
GN	NEUB OR SAG1161.
OS	Streptococcus agalactiae (serotype V).
OC	Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC	Streptococcus.
OX	NCBI_taxid=216466;
[1]	
RN	SEQUENCE FROM N.A.
RP	STRAIN=2603 V/R / Serotype V;
RC	MEDLINE=22222988; PubMed=12200547;
RE	Tettelin H., Maignani V., Cieslewicz M.J., Eisen J.A., Peterson S.,
RA	Wessels M.R., Paulsen I.T., Nelson K.E., Margarit I., Read T.D.,
RA	Madoff L.C., Wolf A.M., Beanan M.J., Brinkac L.M., Daugherty S.C.,
RA	DeBoy R.T., Durkin A.S., Kolonay J.F., Madupu R., Lewis M.R.,
RA	Radune D., Pedorova N.B., Scanlan D., Khouri H., Mulligan S.,
RA	Carthy H.A., Cline R.T., Van Ken S.E., Gill J., Scarselli M., Mora M.,
RA	Iaccobini E.T., Brettoni C., Galli G., Mariani M., Vegni F., Malone D.,
RA	Rinaudo D., Rappuoli R., Telford J.L., Kasper D.L., Grandi G.,
RA	Fraser C.M.;
RT	"Complete genome sequence and comparative genomic analysis of an
RT	emerging human pathogen, serotype V <i>Streptococcus agalactiae</i> .";
EL	Proc. Natl. Acad. Sci. U.S.A. 99:12391-12396 (2002).
DR	EMBL; AE014245; AAN0043.1; -.
DR	PIR; T44651; T44651.
DR	TIGR; SAG1161; -.
DR	GO; GO:0016051; P:carbohydrate biosynthesis; IEA.
DR	InterPro; IPR006014; Antifreeze dom.
DR	InterPro; IPR006190; Antifreeze_like.
DR	InterPro; IPR004144; NeuB.
DR	Pfam; PF01354; Antifreeze; 1.
DR	PFAM; PF03102; NeuB; 1.
DR	PROSITE; PS50844; AFP LIKE; 1.

```

KW Complete proteome.
SQ SEQUENCE 341 AA; 38033 MW; 1488756C2B2BF314 CRC64;

Query Match
Best Local Similarity 36.8%; Pred. No. 1.6e-35; Length 341;
Matches 125; Conservative 66; Mismatches 136; Indels 13; Gaps 6;

Qy 20 FIIAEIQNHQGDLDVAKRMIRMAKCGACAKFQKSELEFKNRKALERPYTSKHSWGK 79
   ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| :
Db 3 YIIAEIGCNHGDIINLAKMVDVAVSGVDVAFQTFKAEKLISKFPAPKAEY-QKATGT 61

Qy 80 TYG--EHRHLEPSSHQYRELQYAEVGIFFTFASGMDENAVFELHNLVPPFKVSGDGT 137
   : || : || : || : || : || : || : || : || : || : || : || : || : ||
Db 62 ADSQLEMTKRLSLFSFEYLEMRDYAISKGVETFTSPFDEESLEFLISTDMPYKIPSGEI 121

Qy 138 NNPPYLEKTAKGKPPMVISSGMSQMDTMKVQYQIVRPLNPN----FCFLOCTSAYPLOPE 193
   ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : |||||
Db 122 TNLPLEKIGKQKKQKVLSTGMA--VMEISIQAVNLRQNGTDISILHCTTEYPTYP 178

Qy 194 DVNLRVISEYQKLPDPIDIPIGSYGHETGIALSVAAVALGAKVLERHITLDTKWGSDHSAS 253
   : || : || : || : || : || : || : || : || : || : || : || : || : ||
Db 179 SLNLNVJHTUKDFEKDUTIGYSDHSTGSEVPDIAAAMGAEVIEKHFTLDTNMGGPDHKAS 238

Qy 254 LEPGELAEVRSVRLVERALGSPTKQLLPCEMACNEKLKGSVVAKVKIPGTTILTMDMLT 313
   ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : |||||
Db 239 ATPDILAALVKGVRIVEQALGRFEKIPDPVEEKNKIVARKSVVALKPIKKGDIYSIENIT 298

Qy 314 VKYGEPI-KAYPPEDIINFLVGKKVLVTVEEDDTIMEELVDN 352

299 VK--RPNGGIGSPMNWYDILGOEADDDFEDEVIRDRSRFN 336

```

```

RESULT 11
Q82UC3
ID Q82UC3 PRELIMINARY; PRT; 749 AA.
AC Q82UC3;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Type III antifreeze protein:CS domain:Neub family (EC 4.1.3.-).
GN NEI570.
OS Nitrosomonas europaea.
OC Bacteria; Proteobacteria; Betaproteobacteria; Nitrosomonadales;
OC Nitrosomonadaceae; Nitrosomonas.
OX NCBI_TaxID=915;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=ATCC 19718 / IF0 14298;
RX MEDLINE=22586410; PubMed=12700255;
RA Chain P., Lamerdin J.E., Larimer F.W., Regala W., Lao V., Land M.,
RA Hauser L., Hooper A.B., Klotz M.G., Norton J., Sayavedra-Soto L.A.,
RA Acierio D.M., Hommes N.G., Whittaker M.M., App D.J.;
RT "complete genome sequence of the ammonia-oxidizing bacterium and
RL J. Bacteriol. 185:2759-2773(2003)".
DR EMBL: BX321861; CAD85481.1; -.
DR GO; GO:0016829; F:lyase activity; IEA.
DR GO; GO:0016051; P:carbohydrate biosynthesis; IEA.
DR InterPro; IPR006014; Antifreeze dom.
DR InterPro; IPR006190; Antifreeze like.
DR InterPro; IPR000644; CS domain.
DR InterPro; IPR004144; NeuB.
DR Pfam; PF01354; Antifreeze; 1.
DR Pfam; PF00571; CS; 1.
DR Pfam; PF03102; Neub; 1.
DR PROSITE; PS50844; APP LIKE; 1.
DR Lyase; Complete proteome.
SQ SEQUENCE 749 AA; 83936 MW; 284BB128108149A9 CRC64;

```

Query Match 27.9%; Score 526; DB 16; Length 749;
Best Local Similarity 35.6%; Pred. No. 1.7e-34;
Matches 115; Conservative 63; Mismatches 139; Indels 6; Gaps 5;

